D

Dairy Cattle Breeding

Filippo Miglior[1,2], Sarah Loker[3], Roger D. Shanks[4]
[1]Guelph Food Research Centre, Agriculture and Agri–Food Canada, ON, Canada
[2]Canadian Dairy Network, Guelph, ON, Canada
[3]Department of Animal and Poultry Science, University of Guelph, Guelph, ON, Canada
[4]Department of Animal Sciences, University of Illinois, Urbana, IL, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Breeding Goals
Data Collection, Identification, and Pedigree Registration
Breeding Scheme
Purebred and Crossbred Cows
Genetic Evaluation
Future Directions
Bibliography

Glossary

Allele A variant form of a gene. Differences between alleles of a gene are a result of alternate DNA sequences.

Breeding value The sum of the independent allele effects on the trait of interest (i.e., the additive genetic worth).

Generation interval The average age of the parent when he/she is replaced by their offspring.

Genome-wide selection Selection of animals based on the value of their genomic profile. Animals are genotyped for several (thousands) of markers spanning the entire genome. These markers are so close together that they are thought to be linked with all genes in the genome.

Heritability A population measure depicting the strength of the relationship between performance and breeding value.

Indicator trait A trait genetically correlated with the trait of interest, but is easier, cheaper, or more convenient to measure and select in hopes of indirectly affecting the trait of interest in the population.

Mendelian sampling Describes the genetic variation of progeny of the same parents. More specifically, full-sibs are not expected to be genetically identical because of random segregation and recombination of genes from the sire and dam.

Reliability Regarding estimated breeding values, reliability, or accuracy of the estimated breeding value reflects the strength of the relationship between the estimated breeding value and the true breeding value.

Definition of the Subject

Dairy cattle breeding is the process of selecting and mating individuals in accordance with breeding goals, with the aim of changing genetic merit of future generations and bringing about an improvement in economic efficiency. For instance, a breeding goal may be designed to improve milk production, health, and fertility. Selection would then be for individuals who will produce offspring that genetically will earn greater profit through improved production at a lower cost (due to improved health and fertility).

Many factors have contributed to the vast improvement in dairy cattle production over the last century. One of the most important factors is the regular recording of phenotypic records. It is from these phenotypic records that the industry has estimated genetic worth. Improvement in methods for genetically evaluating dairy cattle is a large contributor to the

Robert A. Meyers (ed.), *Encyclopedia of Sustainability Science and Technology*, DOI 10.1007/978-1-4419-0851-3,
© Springer Science+Business Media, LLC 2012

substantial genetic improvement seen in this species. Such methods include the use of BLUP (Best Linear Unbiased Prediction), a method first proposed by C.R. Henderson in 1949 [1]. Another important milestone in the improvement of dairy cattle breeding was the development of techniques to freeze and store bovine semen in the early 1950s [2]. Because of this, semen can be stored for longer, shipped further, and therefore shared internationally. A large contributor to the success of dairy cattle breeding has been the implementation of progeny testing programs in the 1950s, which allowed for reliable genetic evaluations for bulls, especially for traits only expressed in daughters (such as milk production) [2]. To aid in the trade and use of dairy cattle genetics on an international basis, Schaeffer in 1994 developed MACE (multiple-trait across-country evaluation) [3]. This methodology allowed genetic evaluations to be converted to different countries' scales. In 2006, Shook [4] described the remarkable increase in yield traits from 1980 to 2000, revealing an increase of 3,500 kg of milk, 130 kg of fat, and 100 kg of protein per cow per lactation. While this increase is due to improvements of many factors, including genetics, nutrition, and management, Shook [4] determined that 55% of the gains in yield traits were due to genetics and that genetic change (versus altering environmental conditions) is permanent and cumulative.

Introduction

As seen in previous entries, genetic improvement in any livestock species requires: (a) identification of breeding goals; (b) accurate data collection, animal identification, and pedigree registration; (c) breeding scheme; and (d) genetic evaluation of measured traits. In dairy cattle breeding, artificial insemination is highly used and traits of interest are usually only expressed in females. Both points determine that males are very important in breeding scheme and genetic progress, but generation interval will be longer than in other species, given that males need to be proven based on progeny performance instead of their own. Another important aspect of dairy cattle breeding is an open international market for dairy genetics, where the male side is controlled through semen sales by a large number of AI organizations, some national and some multinational based. The female side, by contrast, is controlled by the dairy producers. Being an international market with high exchange of semen, and somewhat lower but still common exchange of embryos and live animals, a constant need is to obtain genetic values of foreign animals on local scales, a service provided via international genetic evaluations by the Interbull Centre in Sweden. Finally, in the last 2 years, the full sequence of the bovine genome has opened the way for genome-wide selection. The advent of genomic selection has provided new opportunities and challenges in the global dairy semen market. The market has already seen a partial shift from progeny tested sires to young genotyped bulls. After this transition time, provided one can confirm over the next few months that the genetic level and accuracy of evaluation of these young bulls are as high as expected, genomic selection will revolutionize dairy cattle breeding, and will decrease the importance of progeny testing for some bulls. This chapter will present all the characteristics of traditional dairy cattle breeding, the international aspects of this species breeding as well as the current application of genomic selection and its consequences.

Breeding Goals

Generally, the breeding goal of a dairy producer is to maximize the profitability of his/her dairy farm. The main return in a dairy farm derives from sales of milk production. Cost of milk varies across and within countries based on supply and demand and whether a quota system is present. Additionally, premiums are paid for high-quality milk, and higher percentages of fat and protein. Furthermore, penalization will apply for milk with somatic cell count (SCC) higher than a given threshold. The second return in a dairy farm originates from the sale of breeding stock, primarily young or pregnant heifers. This type of return is less common, being present only in dairy farms with high-genetic-value cows. The same type of farm will generate return from embryo sales by multiple flushings of their top cows. A more common but low return derives from the sale of male calves and cull cows. The most important variable cost in dairy farms is represented by feed costs, followed by veterinary and breeding costs.

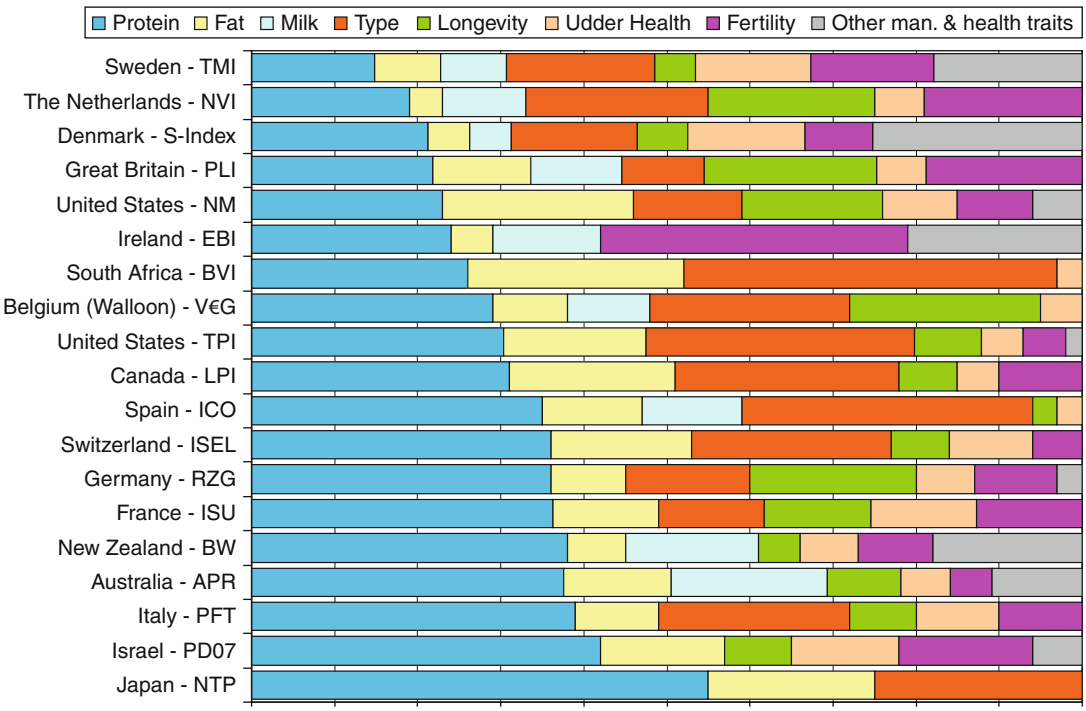
For many years, most selection programs worldwide focused on increasing milk production. National selection indices were based on improving milk yield

and gradually shifted toward improving protein yield and, outside North America, toward increasing fat and especially protein content. This was true for most countries with the exception of the Scandinavian countries, whose selection indices also included health and reproduction; and North American countries, whose selection indices included conformation together with production. In the last 10 years, a growing interest has broadened selection indices to include functional traits such as reproduction and health. Main reasons for this shift were quota-based milk marketing systems, price constraints, or both, together with increasing producer and consumer concerns associated with the observed deterioration of the health and reproduction of dairy cows. Labor costs have increased relatively more than milk price in some countries. Several studies have shown that selection for production alone causes negative effects on udder health [5] and reproductive performance [6–8].

Figure 1 shows the relative emphasis on traits in national selection indices in October 2009. The main difference between selection indices in various countries was the relative emphasis on production. However, every country has now broadened their index by adding longevity, health, and reproduction to the usual production and conformation traits [9]. The search for the ideal balance between all of these important traits continues.

Data Collection, Identification, and Pedigree Registration

Proper identification, pedigree recording, and performance recording are crucial for genetic improvement of dairy cattle. Without them, accurate genetic evaluations would not be possible. In the past, animal identification was more important within the herd for management purposes [10]. However, it is now important to have proper animal identification for genetic evaluation purposes, which means that an animal's ID should be unique outside its herd. In some countries, a unique animal ID is mandatory. In the Canadian dairy cattle industry, a herd's lactation records only qualify for official publication if 80% of its first lactation animals are registered in a breed association herd book with a unique animal identification. Unique



Dairy Cattle Breeding. Figure 1
Relative emphasis of traits for various national selection indexes around the world

animal identification is not without error, however. Larger ID numbers associated with unique national or international identifications are at risk of recording errors, and ID tags can become worn-out or lost. Identification errors can also lead to inaccurate pedigree recording, though pedigree errors can occur for various reasons. Banos et al. [11] and Israel and Weller [12] showed that pedigree errors resulted in biased estimated breeding values and reduced genetic gain. Also, faulty equipment and human error can lead to inaccurate data recording. Fortunately, several techniques correct or accommodate erroneous outliers in the data (e.g., robust procedures described by Jamrozik et al. [13]).

In summary, numerous errors can occur when recording performance and pedigree information. Therefore, to ensure better quality data, it is mandatory in most countries to follow the rules and standards established by the International Committee for Animal Recording before records can be used for genetic evaluation. Several traits are of economic importance in the dairy industry, but the relative importance of each depends on the country (Fig. 1).

Test-day models are used for the genetic evaluation of dairy cattle for milk production traits in many countries. These models necessitate the regular recording of milk production traits. A good recording scheme can therefore require records for 24-h milk, fat, protein, and somatic cell count (SCC) to be taken once monthly. These are called test-day records. Generally, production traits such as milk, fat, and protein yield are moderately heritable.

While milk production traits are important, many other traits are of economic importance in dairy cattle breeding, including conformation, longevity, reproduction, health, and workability traits. Conformation (or "type") traits describe the physical attributes of the cow that are generally associated with survival, health, and reproduction. Many traits (e.g., body condition score) require visual appraisal by the recorder, and are considered to be more subjective. In these cases, it is vital that assessors are highly trained to ensure repeatable and accurate recording. Many "type" traits are moderately heritable.

Because of the negative genetic correlations between milk production and fertility or health traits, long-term selection for improved milk production has led to reduced fertility and health in dairy cattle. As a result, routine genetic evaluations of reproductive and health traits are becoming more common, despite low heritabilities. A major challenge is that direct health data has not been recorded for very long and can be difficult to measure. In many cases, countries use indicator traits instead of measuring the health trait directly. An example of this is using somatic cell count as an indicator of mastitis.

Longevity (or survival) describes the length of a cow's survival in the herd, and has a low heritability as this trait can be greatly affected by herd management and other nongenetic factors. Workability includes traits such as milking speed and temperament during milking. Milking temperament has a low heritability, while milking speed has a moderate heritability.

Breeding Scheme

In general, a breeding scheme is the amalgamation of the processes involved in the selection and mating of livestock for the purpose of genetic improvement. Because of artificial insemination in the dairy cattle industry, semen from a single male can be used widely throughout the population. Therefore, genetic improvement is achieved largely through intense selection of males. However, most of the economically important traits in the industry (such as milk production traits) are expressed in the female. As a result, dairy cattle breeding relies on progeny testing schemes for genetic improvement. Data on various milk production and performance traits from daughters are collected and used to calculate estimated breeding values for bulls. The more daughters and daughter records are available for a bull, the greater the accuracy of the estimated breeding values. "Proven bulls" are bulls with very reliable estimated breeding values because they have many daughters with performance records. Estimated breeding values for bulls without progeny records are less reliable because they are calculated using the average of the estimated breeding values of the parents.

The major factors influencing rate of genetic progress for a given trait are the components of "the key equation" of animal breeding:

$$\frac{\Delta BV}{t} = \frac{r_{BV,\,\widehat{BV}}\, i\sigma_{BV}}{L} \qquad (1)$$

where BV is true breeding value, t is time, $\widehat{BV}$= estimated breeding value, $r_{BV,\widehat{BV}}$= accuracy of the estimated breeding value relative to the true breeding value (also the correlation between the estimated and true breeding values), i = selection intensity (a function of the proportion of the population chosen to be parents of the next generation), σ_{BV} is additive genetic standard deviation, and L is generation interval. The aim is to choose animals with superior genetics to be parents of the next generation to improve the genetics of the population. Increasing the reliability (accuracy) of prediction, selecting only the best animals as parents (increasing selection intensity), and decreasing the generation interval are important factors for increasing genetic improvement per unit time.

A major challenge in dairy cattle breeding is developing an optimum breeding program that maximizes genetic progress while minimizing cost. AI organizations are generally responsible for breeding schemes [14], and a typical AI organization can spend millions of dollars per year progeny testing bulls to find the best bull to market to the world [15]. However, dairy breeding is entering a new era in which genomic selection is possible. With genomic selection, bulls can be genotyped and selected at a young age. This improves the industry's traditional breeding scheme by reducing reliance on progeny testing (a lengthy and costly process), and will theoretically increase response to selection via a reduced generation interval [15]. Also, genomic values will increase the accuracy of genetic evaluations, especially for young sires for which traditional estimated breeding values are derived from parent averages [16].

Purebred and Crossbred Cows

The group of animals selected as parents of the next generation are expected to possess alleles that the industry considers favorable. Therefore, through selection, the frequency of favorable alleles in the population should increase with each generation while that of unfavorable alleles should decrease. The result is an increase in average breeding value, and improved performance of the dairy cattle population. The change in average breeding value over time defines the genetic trend.

As mentioned previously, artificial insemination has allowed for intense selection of sires for increased genetic improvement over time. This means that a few top bulls, with the best collection of favorable alleles, can be mated widely throughout the population. While this is a good way to progress more quickly toward fixing favorable alleles in the population, it reduces the effective population size of the breed which could raise inbreeding and reduce performance from the associated inbreeding depression. The dairy industry needs to find a compromise between selection of the best sires for use in artificial insemination, and minimization of inbreeding depression. This is, of course, less of a problem initially with crossbreeding, as sire and dam are unrelated. However, while several benefits exist with crossbreeding in general, heterosis obtained is too low to lead to more profitable animals than purebred Holsteins (the most widely used dairy breed) [17, 18]. Therefore, in the dairy industry, selection tends to occur within breeds.

Again, a large degree of dairy cattle genetic progress is achieved through the selection and use of semen of a few top bulls. However, it is important to understand that selection and consequent genetic progress within a breed is achieved via four selection pathways, all of which center around the progeny testing scheme.

Progeny testing is required so that the genetic merit of bulls can be calculated reliably via analysis of many performance records on many daughters. Every year, genetically superior bulls and cows are mated using artificial insemination to produce young bulls with high predicted genetic merit. On average, young bulls resulting from these matings will have high true genetic merit, but because of Mendelian sampling, it is not certain that these young bulls will be genetically superior. The young bulls therefore need to be proven via progeny testing. If a young bull is in fact of high genetic merit, he will yield daughters that perform well for traits of interest. The more daughters he produces with superior performance, the more certain it is that he is a genetically superior bull.

So far, two selection pathways have been discussed: selection of sires of young bulls, and of dams of young bulls. The sires of young bulls are proven and their semen can be used widely, so they can be selected very intensely, and their estimated breeding values are very reliable. Referring to Eq. 1, increased selection pressure and reliability will lead to increased genetic progress per unit time. Dams of young bulls can be selected

intensely because not many young bulls need to be produced for progeny testing (only about 400 a year in Canada and about 6,000 Holsteins worldwide). However, the reliability of estimated breeding values for dams of young bulls is not as high as that of the sires because the dams have fewer close relatives with records.

In a dairy herd, replacement heifers are required. Therefore, further potential for increasing genetic progress of the population is through two more selection pathways: selection of sires of cows and dams of cows. It takes several years for young bulls to be fully proven. However, when these bulls have a genomic evaluation or some daughter records (but not yet enough to achieve the reliability of a proven bull), they can be selected to produce replacement heifers with a reasonably high selection intensity. Dams of these future cows, however, cannot be selected so intensely for several reasons. Many replacement heifers are required, and because a female cannot breed as many times as a male, most cows will be selected for breeding. Also, female fertility in the dairy cattle population is typically low, so the industry cannot afford to be very selective with this pathway.

Genetic Evaluation

For progress in the dairy industry, it is important to accurately select genetically superior animals as parents of the next generation. Traditionally, genetic worth could only be estimated by evaluation of phenotypic records, which are a result of a combination of genetic and environmental factors (and sometimes an interaction of the two). Again, genomics is revolutionizing the way the industry evaluates dairy cattle, making it possible to genotype animals instead of waiting for phenotypic records. However, genomics is just one part of the process, and the collection of phenotypic records will still be important for some time.

The additive genetic value of an animal for a particular trait is the sum of the independent effects of that individual's alleles on that trait. On average, half of an animal's additive genetic worth is passed on to its offspring. The greater the animal's genetic worth, the more genetically superior its offspring are expected to be. The additive genetic value is therefore appropriately termed "breeding value." Breeding values of animals can be estimated from many different sources of information, including observations on the animal itself and observations from a variety of relatives. This reiterates the importance of quality phenotypic and pedigree data. True breeding value can never be known, only estimated from a very large (effectively infinite) number of genes with alleles each of which has a small effect on the trait of interest. Breeding values are estimated from limited phenotypic data using models that are not perfect. The accuracy of estimated breeding values (EBV) depends on a variety of factors, including the degree of relationship between the animals providing the phenotypic information and the animal being evaluated, the number of records available, and the heritability of the trait of interest. Of course, through genomics, perhaps one day the effect that each allele in the genome has on each trait could be quantified, bringing the industry closer to an animal's true breeding value for each trait.

Several traits of economic importance to the dairy industry were already discussed. So, a dairy bull or cow has several estimated breeding values, one for each trait. This makes selection complicated. For example, perhaps a potential sire has excellent estimated breeding values for milk production traits, but terrible values for health and fertility traits. Therefore, countries devise a national economic selection index, which incorporates estimated breeding values for traits of interest and their respective monetary worth into an equation that gives a single score for profitability ("aggregate breeding value") of each animal [19]. This makes selection much easier, as animals with the most favorable combination of estimated breeding values (e.g., high milk production and good health and fertility) are the most profitable.

Traditionally, selection index methods were used to combine weighting factors with adjusted phenotypic records from various sources (i.e., own records and records from various relatives) to derive estimated breeding values for traits. The phenotypic records were first adjusted for a variety of environmental influences including, for example, age of animal and effect of contemporary group. These methods of adjustment were not always effective for disentangling genetic effects from environmental influences. To improve the application of selection index methods, the dairy breeding industry began to use the statistical method

known as best linear unbiased prediction, or BLUP, in the 1970s. This method was first proposed by C.R. Henderson in 1949 [1]. Without going into too much detail, BLUP is able to simultaneously estimate environmental effects and predict breeding values while taking into account pedigree relationships.

Both the traits and the methodologies involved in national genetic evaluations vary substantially among countries [20]. Therefore, EBVs for one trait in one country may not be representative of EBVs for the same trait in another country. This makes comparing animals from different countries difficult. Dairy cattle genetics are shared internationally, especially sire genetics via artificial insemination. Therefore, the Interbull Centre was created to provide international evaluations. Specifically, the procedure carried out is called the multiple-trait across-country evaluation, or MACE [3]. This procedure allows Interbull to provide a separate list of International Genetic Evaluations to each participating country, expressed on that country's scale.

Future Directions

As previously mentioned, genome-wide selection is revolutionizing dairy cattle breeding. Young bulls benefit the most with a large increase in reliability of estimated breeding values at an earlier time in their lives, reducing the generation interval of these animals and hence increasing the speed of genetic improvement. It is fairly certain that future dairy cattle genetics research will focus on the improvement of genomic techniques.

Over the years, as quantitative geneticists improve upon techniques surrounding genome-wide selection for animal breeding, it is important to keep in mind the application of such techniques to human health. In 2008, Mardis [21] predicted that sequencing the entire human genome for $1,000 will be feasible in the near future. While animal breeders are currently using genomics to predict the genetic value of animals for complex traits, it may one day be possible to utilize genomics to predict human individuals' genetic risk for complex, multifactorial diseases, such as Crohn's disease [22]. Research in genomics in animal breeding will certainly pave the way for research and development of genomics applied to human health.

Bibliography

1. Henderson CR (1949) Estimation of changes in herd environment. J Dairy Sci 32:706 (Abstr)
2. Funk DA (2006) Major advances in globalization and consolidation of the artificial insemination industry. J Dairy Sci 89:1362–1368
3. Schaeffer LR (1994) Multiple-country comparison of dairy sires. J Dairy Sci 77:2671–2678
4. Shook GE (2006) Major advances in determining appropriate selection goals. J Dairy Sci 89:1349–1361
5. Heringstad B, Chang YM, Gianola D, Klemetsdal G (2003) Genetic analysis of longitudinal trajectory of clinical mastitis in first-lactation Norwegian cattle. J Dairy Sci 86: 2676–2683
6. Haile-Mariam M, Morton JM, Goddard ME (2003) Estimates of genetic parameters for fertility traits of Australian Holstein-Friesian cattle. Anim Sci 76:35–42
7. Kadarmideen HN, Thompson R, Coffey MP, Kossaibati MA (2003) Genetic parameters and evaluations from single- and multiple- trait analysis of dairy cow fertility and milk production. Livest Prod Sci 81:183–195
8. Veerkamp RF, Koenen EPC, De Jong G (2001) Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models. J Dairy Sci 84:2327–2335
9. Miglior F, Muir BL, Van Doormaal BJ (2005) Selection indices in Holstein cattle of various countries. J Dairy Sci 88:1255–1263
10. Jr Hooven NW (1978) Cow identification and recording systems. J Dairy Sci 61:1167–1180
11. Banos G, Wiggans GR, Powell RL (2001) Impact of paternity errors in cow identification on genetic evaluations and international comparisons. J Dairy Sci 84:2523–2529
12. Israel C, Weller JI (2000) Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J Dairy Sci 83:181–187
13. Jamrozik J, Strandén I, Schaeffer LR (2004) Random regression test-day models with residuals following a Student's-*t* distribution. J Dairy Sci 87:699–705
14. Miglior F (2000) Impact of inbreeding – Managing a declining Holstein gene pool. In: Proceedings from 10th world Holstein Friesian federation conference – 30 Apr – 3 May 2000, Sydney, pp 108–113
15. Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223
16. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24
17. Gunjal K, Menard L, Shanmugam R (1997) Economic analysis of crossbreeding dairy cattle. Agric Syst 54:327–339
18. VanRaden PM, Sanders AH (2003) Economic merit of crossbred and purebred US dairy cattle. J Dairy Sci 86:1036–1044
19. Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 28:476–490

20. Mark T (2004) Applied genetic evaluations for production and functional traits in dairy cattle. J Dairy Sci 87:2641–2652
21. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402
22. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH et al (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40:955–962

Dam Engineering and its Environmental Aspects

Petar T. Milanović
Belgrade, Serbia

Article Outline

Glossary
Definition of the Subject
Introduction
Flood Regulation: Regional Impact on Population
Dam Failures
Reservoir Slope Instability
Tailings Dam Failures
Reservoir-Triggered Seismicity: Induced Seismicity
Dams and Heritage Protection
Dams and Ecosystems
Dams and Microclimate
Induced Subsidence (Collapse)
Spring Submergence Due to Dam Construction
Environmental Aspects of Dams in Karst
Further Directions
Bibliography

Glossary

Dam Civil structure planned, constructed, and operated to meet human needs in flood control, irrigation, supply of drinking water, electricity generation, recreation, and various other purposes.

Dam failure Collapse or movement of part of a dam or its foundation, so that the dam cannot retain water.

Guaranty ecological flow Required quantity and quality of flow to maintain the sustainability of the river ecosystem (ecological base flow).

Induced subsidence Collapse of the surface of the ground due to human activities, mostly reservoir operation and intensive pumping of groundwater.

Karst Terrain composed of highly soluble rocks (limestone, dolomite, gypsum, and salt), very risky environment for dams and reservoirs construction.

Large dams Dams having a height of 15 m from the foundation or, if the height is between 5 and 15 m, having a reservoir capacity of more than 3 million cubic meters.

Reservoir-triggered seismicity Seismic phenomena associated with impounding of reservoirs (Reservoir-Induced Seismicity).

Definition of the Subject

Construction of dams and reservoirs involves considerable natural and anthropogenic impacts. These impacts are for the most part positive, but can some negative influence on the environment. The main purposes of dam construction are focused on water regime improvement, and consequently to regional prosperity. Generally, the goal of dams and reservoirs is regional socioeconomic development by irrigation, flood control, power production, water supply, recreation purposes, reduction of deforestation, reduction of drought periods, for fishing farms, mining purposes, navigation, to enhance landscape including development of new infrastructure, and to provide new possibility for employment and many secondary benefits.

However, as a consequence of dam and reservoirs construction, a number of different and sometimes unpredictable negative environmental impacts and uncertainties cannot be avoided. Some common negative impacts are: the population migrates from inundated areas; the reservoirs cover arable land, settlements, and infrastructure; deep reservoirs provoke induced seismicity and induced collapses; water fluctuation provokes landslides along the reservoir banks; sedimentation of reservoirs; in some cases, important cultural and historical monuments are inundated; questionable impact on biodiversities, survival of wildlife, and endemic species is endangered; tailings may contain dangerous chemicals; and regime of surface and underground water is considerably changed. In a number of cases, socioeconomic constrains related to migration from submerged regions are very pronounced.

The worst most negative and disastrous impact is dam failures. The karst environment is a particularly sensitive influence on dams and reservoirs, and a variety of positive and negative consequences are numerous in such terrain.

Introduction

Since time immemorial, people have considered how to tame the surface waters to prevent floods and to use water for different purposes, irrigation, water supply, and much later, for electricity production. In many areas in the world, life was not determined by man, but by the rivers. The people have had to cope with two kinds of misfortune: flood and drought. In many cases, the consequences were disastrous. One single flood of the Yangtze River in China (1931) devastated 3.3 million hectares of arable land and caused the suffering of 28 million people and the loss of more than 145,000 lives [1].

From very ancient times, dams appeared as the only effective structures to tame river waters. Construction of dams started a few 1,000 years ago. Primary role of dams is to store or to divert waters. The oldest known are Jawa dams in Jordan (~ 3000 BC). In Egypt, the Kosheish Dam was constructed during the period 3000–2900 BC and Saad El-Kafra Dam about 2610 BC. The Anfantang reservoir in China was built in sixth century BC, and a 30-m high gabion dam was constructed around 240 BC in Shanxi province. In Iran, dam constructions dates from before 2,000 years ago (Bahman Dam, Fig. 1); Shapour and Mizan dams were constructed during the reign of King Shapur I about 1,700 years ago; the Tilkan and Sheshtarz Dam, 1,000 years ago. The Amir Dam north of Shiraz, 1,000 years old, still is operational [2]. In Spain, the Proserpina Dam, 22 m high, was built in the second century and still is operational.

Later in the twentieth century, individual dams were constructed to control the large hydro-systems, which consist of a number of dams and reservoirs, to change water regime at large catchment areas: Tennessee Valley Authority (29 dams) USA; dams along the Yangtze River in China; Southeast Anatolia Project (21 dams); the Volga-Kama cascades, Russia (11 dams); dams at Karun River catchment in Iran (16 dams); or

Dam Engineering and its Environmental Aspects. Figure 1
Bahman Dam, Iran. Constructed approximately 2,200 years ago

dams in karst area the Hydro-system Trebišnjica, Bosnia and Herzegovina (BiH), and Croatia (6 dams).

In many cases, dam projects were initially controversial and potentially disastrous. Failures of dams and potential environmental impacts are the main reasons for controversy and fear. However, with increasing demand on water resources and electric power, at issue is how to keep the balance between the necessity for development and preservation of the environment. At many locations, dams are still structures of great importance.

According ICOLD World Register of Large Dams (1998), the main purpose of large dams is irrigation – 37%, multipurpose use – 22%, electricity generation – 16%, water supply – 12%, flood control – 6%, recreation – 3%, and other purposes – 4%. Tailing dams are excluded from this survey.

In the USA, considering all dams (not large dams only), the main purpose is recreation – 33.8%, flood control – 15.6%, fire control 13.7%, irrigation – 9.5%, water supply – 9.4%, electricity production – 2.9%, and rest for many different purposes.

Environmental problems are coupled with political (transboundary) problems if dams are constructed at rivers bordering countries: for example, the Iron Gate Dam at the Danube River between Serbia and Romania; the Aswan Reservoir at the Nile River between Egypt and Sudan; dams on the Euphrates and Tigris Rivers impacting Turkey, Syria, and Iraq; and the Itaipu Reservoir along the border between Brazil and Paraguay.

Flood Regulation: Regional Impact on Population

Historically, for thousands of years the primary role of dams was protection of settlements and arable land against flooding and for irrigation. Presently, flood control and irrigation are still of high importance, but usually dams and reservoirs are multifunctional, including power production, water supply, sediment control, landscape improvement, and recreation.

The Tennessee Valley Authority (Tennessee River, USA), founded in 1933, is one of the largest dam reservoir projects for flood control, navigation, power production, and irrigation [3]. Under natural conditions, the water regime was unfavorable for agriculture and life. Thirty percent of the population in the Tennessee Valley was affected by malaria. To construct 29 dams and reservoirs, more than 15,000 families were displaced. Electricity generation, flood control, and better organized water regimes have been of great benefit for the region. More than 1,000 km of navigation channels have been constructed as part of this project. One important positive environmental impact is control of surface waters with regard to malaria prevention.

Large floods were a regular occurrence along the Nile River under natural conditions. At the same time, floodwater deposited about 4 million tons of nutrient-rich sediments per year. To prevent floods, the first modern dam at Aswan was built in 1889. The first dam was not high or effective enough to control flooding, and a new Aswan High Dam was constructed during 1960–1970 a few kilometers upstream. Over 60,000 Nubians were relocated from the reservoir area. The reservoir contains a volume of 169 billion cubic meters. About 17% of the reservoir is in Sudan. After dam construction, the annual floods are under control and the navigation properties of the Nile River are considerably improved. However, artificial fertilizers have to be used instead of natural nutrients. Quality of the soil for farming has decreased. Negative impacts due to lack of rich sediments in the delta region is one negative environmental consequence.

Over the past two centuries, many people have died along the Yangtze River, China due to catastrophic floods. In 1840, about 156,000 persons lost their lives during flood periods; in 1931, 145,000; in 1954 about 33,000; and during more recent flooding in 1998, over 1,500 people died. Millions of hectares of arable land has been destroyed and is unusable. Numerous villages have completely disappeared. During the flood of 1954, 18 million were forced to evacuate from the area. For 3 months, Wuhan City with eight million people was covered with floodwater. After construction of the Three Gorges Dam (181 m high, 2,335 m long) the frequency of major floods has been reduced to a minimum and after project completion, large ships will be able to navigate from Shanghai 2,400 km upstream. Due to project requirements, about 1.24 million people had to be relocated.

As result of construction of the Chira-Piura irrigation system in Peru, including the 9-km long Poechos Dam, an area of 100,000 ha of uncultivated land has been turned into fertile farmland.

Construction of the Akosombo Dam in Ghana created one of the largest world reservoirs, Lake Volta; about 15,000 houses were inundated and 78,000 people were resettled.

The Gibe III Dam (240 m high at Omo River) in Ethiopia was controversial from its beginning. During the Omo River flood of 2006, at least 360 people and thousands of livestock were devastated. From an energy viewpoint, this project would provide great benefits for Ethiopia and Kenya. However, equally important is the projected negative impact on fisheries in adjoining Kenya's Lake Turkana, which threatens the livelihoods of about 0.5 million tribal people.

To construct the Ataturk Reservoir in Turkey, about 55,000 people were relocated, and from the Manantali Reservoir area (Mali) 15,000 people were displaced. It is estimated that the controversial Ilisu Dam Project (Turkey) would displace at least 55,000 people. For construction of the Everkiiskaya Dam (Central Siberia, Russia), about 7,000 local indigenous people would need to be relocated and a huge area flooded (9,000 km^2).

Dam Failures

One of the worst negative environmental impacts of dams is the risk of failure. The first known dam failure, as reported by Herodotus, was the collapse of the Saad Ei-Kafra Dam in Egypt during flooding ~2500 BC. In modern history, one of the oldest reported failures was the Blackbrook I Dam (Great Britain, 1799). Official worldwide database and case histories of dam incidents and failures are not still completed. According to the ICOLD Bulletin 99, Dam Failures, Statistical Analysis (without China), a total of 5,268 dams were built until 1950 (117 of them failed), 12,138 dams were built during the years 1951–1986, (59 of them failed).

Many incidents of dam failure occurred more than 50 years ago and involved older dams. In recent decades, the failure rate (particularly for dams more than 30 m high) has drastically decreased. At present, dams are constructed on the basis of thorough and detailed multidisciplinary investigations, including application of new technologies and detailed environmental studies and operational stability is regularly monitored.

The world's recorded dam disaster occurred 1975 in China (Banqiao Dam). Due to a strong hurricane and precipitation of 500 m over 3 days, the Banqiao Dam, together with Shimantan Dam and 62 small dams, was totally demolished. Water waves as high as 6–10 m and 12 km wide flooded more than a million hectares. More than 26,000 people were killed, and many more died afterward from resulting epidemics, for an estimated 150,000 total deaths. Eleven million people lost their homes.

Dam failures as a consequence of geology are sometimes catastrophic, causing loss of life or evacuation of thousands of people living downstream of the dams: for example, Malpasset (France), St. Francisco and St. Fernando (USA), Baldwin Hills (USA), and Teton (USA).

Total failure of the Malpasset concrete arch dam (66.5 m high) caused a huge flood on December 1959. Dam foundation consisted of gneissic rocks. Two sets of faults had crucial roles in the creation of a wedge failure. More than 325 persons lost their lives and a large area was devastated [4].

The worst American civil engineering failure of the twentieth century was the St. Francisco Dam (California, USA), killing 450 people along the St. Francisco Canyon and St. Clara Valley. Foundation rock of the concrete gravity dam (60 m high) is conglomerate. During March 12–13, 1928, the St. Francisco Dam collapsed due to a paleo-landslide at the left abutment and strong uplift. According to Cooper and Calow (1998), the failure was partially attributed to gypsum dissolution [5].

The Teton Dam (USA) failure happened on June 5, 1976, killing 14 people; that embankment dam was 93 m high above riverbed. Dam foundation consists of basalt and rhyolite rock. Piping was identified as the most probable cause of failure [4].

Catastrophic failure of San Juan earth dam (Spain) occurred during the first filling of reservoir (2001). Due to intensive dissolution of gypsum, part of dam collapsed, provoking a huge flood downstream [6]. The flood caused by failure of the Lower San Fernando Dam (California, USA, 1971) due to a strong earthquake (M = 6.6) caused temporary evacuation of 80,000 people from the downstream area [4]. The Baldwin Hills Reservoir (USA) failed in 1963 causing enormous damage downstream; however, thousands of inhabitants were evacuated in time [4].

Thousands of inhabitants downstream from dams all over the world are permanently under psychological

and mental pressure due to possible dam failure. Some of them seek relocation to safer places. If an area downstream from dam is populated, flood wave analysis in case of dam failure is very important. Analysis includes timing of water wave propagation, estimation of flood level, and installation of emergency alarm systems. Two important parameters are: time of flood wave between dam site and urban areas, and safe elevation for population evacuation – the elevation above which a disastrous wave is not effective.

In many cases, citizens living downstream from dams protest strongly against them, sometimes insisting on dam displacement. In the case of the Mulholland Dam (65-m high conrete gravity dam) because of disaster of St. Francisco, there was enormous protest from the citisens of Holliwood lining downstream from the dam. To solve this mostly psychological problem, the downstream dam face was covered in 1933 by a huge earthen mass, making it one of the most conservative dam structures in the world.

Reservoir Slope Instability

Reservoir slopes are exposed to different kinds of hazards. The most common is the potential for landslides to cause a wave which might overtop the dam crest, causing dam failure and disastrous flood wave downstream. According to Schuster (2006), at least 254 large dams worldwide have been subjected to landslide activity [7]. The most common types of hazard are instability of slopes and deterioration of reservoir water quality due to solution processes if the slope consists of evaporates. The most frequent remedial measures to prevent instable rock masses from sliding are retention walls, prestress anchors, galleries and other drainage structures, and grouting and cutting of sliding planes.

Due to hydrodynamic force caused by reservoir fluctuation, the slopes are exposed to sliding, creeping, and toppling. This process can have catastrophic consequences. The difficulty is how to predict potential for a landslide on the basis of geological data and geological history of reservoir rims. An attempt has been presented by Moon (1997) in New Zealand [8]. He has established a magnitude-frequency curve for landslides based on geomorphological evidence of sliding in a valley and the geomorphological history of a valley.

In the case of Vaiont Dam (261-m high dam in Italy), a huge landslide suddenly slid into the reservoir on October 9, 1963. The event lasted only 45 s, but a volume of about 300,000 million cubic meters plunged into the reservoir. The landslide length was 1.850 km and average thickness was 157 m. Maximal thickness of the slide rock mass was 330 m. A water wave was created which overtopped the 100-m high dam crest. The catastrophic water wave completely demolished the small town of Longarone, 2 km downstream, and six more settlements. About 1,700 inhabitants lost their lives and a number of industrial structures were completely destroyed [9]. The dam structure itself was not damaged at all.

In a number of cases, massive stabilization structures are constructed to eliminate hazards during reservoir operation. Induced slope stability is a problem along the 650 km of the Three Gorges Reservoir. The Lianziya potentially sliding rock mass, located about 25 km from the upper stream of the Sadoung dam site of the Three Gorges (China), has a volume of 2.26 million cubic meters. Deep prestress bolting, up to 3,000 kN, is used to improve slope stability [10].

Tailings Dam Failures

Tailings dams are more vulnerable than other dam types. In the case of failure, environmental impact is catastrophic and long lasting. Tailings usually contain high concentrations of different chemicals. They represent a potential threat of environmental contamination, in some cases by extremely dangerous chemicals such as heavy metals or cyanides. One of the latest incidents (Baia Mare, Romania) occurred in January 2000, and released about 100,000 m^3 cyanide contaminated water into the catchment area of Tisa River (tributary of Danube River), provoking great public concern in the huge and highly populated downstream area.

The last "Chronology of major tailings dam failure, 1960–2011" is prepared on the basis of Bulletin 121, published by United Nations Environmental Programme (UNDP), Division of Technology, Industry and Economics (DTIE), and International Commission on Large Dams (ICOLD), Paris 2001, 144 p., updated, March, 2011, (221 tailings dam incidents). Tailings dam failures occurred for many different

reasons, but mostly after heavy rainfall due to overtopping, seepage, foundation failure, or due to dam wall failure or liquefaction during earthquakes. Poor management and inadequate construction methods can also contribute to dam failure [11].

In many cases, failures are disastrous. Heavily polluted tailings flow can travel from a few 100 m up to more than 100 km. After failure of the tailing dam at Silverton, Colorado, USA (1975), tailings flow polluted 160 km of the Animas River; near Fort Mead, Florida, USA (1971), tailings flow traveled 120 km; in the Inez failure in Martin County, Kentucky, USA (2000), polluted flow traveled 120 km; in the case of Huancavelica, Peru (2010), the Escalera and Opamayo Rivers were contaminated 110 km downstream; and during dam failure of El Pocho, Bolivia (1996), 300 km of the Pilcomayo River were contaminated.

The worst impacts of tailings dam failure are the great number of people killed. From 1960 to 2010, 23 cases of tailings failures with fatalities were reported. Drastic examples are failure of the Sgorigrad, 1966 (Bulgaria), 488 people killed; the Stava, 1985 (Trento, Italy), 268 killed; Taoshi, 2008 (Shanxi province, China), 254 killed; El Cordobe, 1965 (Chile), 200 killed; Aberfan, 1966 (Wels, UK), 144 killed; Buffalo Creek, 1972 (W. Virginia, USA), 125 killed; and Mufulira 1970 (Zambia), 89 killed. Common impacts after tailing failures are demolition of homes, relocation of people, inundation of agricultural land, and catastrophic consequences for biodiversities in downstream areas, particularly for fish.

According Rico et al. [12], for a group of 147 cases of worldwide tailings failures, 39% happened in the USA, 12% in Chile, 10% in the UK, and 4.8% in the Philippines [12]. Twenty-six occurred in Europe. With regard to tailings dam height, a greater percentage of failures occur if the height is not higher than 30 m. Tailings failures are frequently related to heavy precipitation or due to seismic liquefaction. More than 85% of failures occurred during mine operation, and only 15% of incidents were related to inactive tailings dams.

Reservoir-Triggered Seismicity: Induced Seismicity

From the very beginning, reservoir-triggered seismicity has been controversial. The first documented case was the case of the Hoover Dam (Lake Mead, USA). Presently, more than 60 cases of reservoirs are frequently cited to have experienced reservoir-triggered seismicity (Perman et al. 1983 and Gupta 1992). Magnitudes are greater in the case of greater depths of reservoirs: Koyna (India), depth 100 m, M = 6.3; Kremasta (Greece), 120 m, M = 6.3; Kariba (Zambia), 122 m, M = 6.25; Hsingfengiang (China), 105 m, M = 6.0; Srinagarind (Thailand), 133 m, M = 5.9; Oroville (USA), 204 m, M = 5.8; Aswan (Egypt), 90 m, M = 5.2; Hoover (USA), 191 m, M = 5.0; Kurobe (Japan), 186 m, M = 4.9; and Mratinje (Montenegro), 220 m, M = 4.1. However, except for a few cases (Koyna and Hsingfengiang), human and material losses were negligible. The Hsingfengiang Dam was considerably damaged, but so far no dam has collapsed due to the effect of induced seismicity [13].

In general, triggered seismicity starts during the first impounding of the reservoir and increases with reservoir water levels; intensity of shaking sharply decreases with distance from the reservoir.

Certain earthquakes registered during reservoir impounding in karst indicate the role of karst in genesis of induced shocks. Those analyses indicate possible explosions of the compressed air during an abrupt reservoir impounding and simultaneous abrupt rising of the water table in the surrounding karst aquifer. Pressure of the air trapped in karst channels and siphons significantly increases. Trapped air pillows escape, creating strong explosions that are felt by inhabitants and recorded by seismographs at the surface. Environmental impact of this process is generally local, noisy, but not harmful. Small damage is possible in the case of older village structures.

Dams and Heritage Protection

During dam construction, in some cases, very important monuments and internationally recognized old civilization heritage sites are threatened by inundations. Some are world heritage sites protected by UNESCO. In many cases, reservoirs may flood national parks, caves of archeological importance, or old necropoli, monasteries, graveyards, and ancient bridges.

The most famous are the Abu Simbel temples in Egypt built in the middle of the Nubian Desert (present

reservoir) by Ramses II who ruled from 1290 to 1224 BC (19th dynasty). After construction of the Aswan Dam, both monuments of Ramses II and of his wife were sawn into 1,036 blocks, 30 t each, plus 1,110 blocks from surrounding rock. Monuments were reconstructed 90 m above the original level (Fig. 2).

Due to construction of dams along the Trebišnjica River (BiH), two ancient monasteries built in 1232 and during the first part of the fourteenth century, and one bridge constructed at the first part of the sixteenth century, nationally recognized cultural heritage sites have been relocated from the reservoir areas (Fig. 3).

Construction of the Lower Gordon dam in southwest Tasmania would have flooded a large karst area containing caves of great archeological importance. The project was abandoned for legal and environmental reasons in 1983 [14].

During construction of the Iron Gate Dam on the Danube River, a number of historical monuments, including ruins of the old bridge over the Danube River dating from 28 to 104 AD (time of Roman Emperors Tiberius, Claudius, and Traianus) have been flooded. To preserve the monument "Tabula Traiana," heavy limestone blocks of 250 t were sawn and lifted 20 m to be above the reservoir level. The old prehistoric settlement (Lepenski Vir) dated between five and six millennia BC, which represents one of the oldest cultures in this part of Europe, was relocated above the Danube Reservoir level.

Along the Three Gorges Reservoir (Yangtze River, China) about 1,300 archeological sites, including 30 Stone Age localities, have been carefully investigated. About 1,200 of them will be relocated to higher places. Some irreplaceable historical artifacts, however, have been permanently inundated.

The old Greek and Roman City of Zeugma, larger than Pompeii, founded in 300 BC, on the Euphrates River, was inundated after construction of the Birecik Dam, Southeast Anadolian Project, Turkey (1999). Zeugma mosaics have been declared one of the best preserved Roman Mosaic collections in the world. This ancient city is submerged, but its famous mosaics were placed in the museum of Gazi Antep.

The proposed Ilisu Dam (Southeast Anadolian Project, Turkey) upper Tigris River, 65 km from the Syrian border, threatens to inundate Hasankeyf, an

Dam Engineering and its Environmental Aspects. Figure 2
Abu Simbel monument replaced from Aswan Reservoir, Egypt

Dam Engineering and its Environmental Aspects. Figure 3
Old monasteries at bottom of Bileća Reservoir (BiH)

internationally recognized Roman, Byzantine, and Ottoman historical and cultural heritage site. Historical monuments include thousands of caves carved more than 2,000 years ago. About 50 villages and 15 small towns along the Tigris valley would be displaced [15].

A large part of Munzur National Park (Turkey) is to be flooded by construction of eight dams. Thousands of endemic species will become extinct after finalization of the project. Some of these dams are already constructed (Mercan and Uzuncayir dams).

A dam project in Coa Valley, Portugal was canceled because of the important Ice Age rock art. Many irreplaceable artifacts of an ancient Mesopotamian city (2000 BC) are endangered by construction of the Makhal Dam in Iraq. In the case of construction of Sardar Sarovar and Narmada Sagar dams at Narmada River (India) 250,000 people will be displaced and 3,000-year-old historical temples are potentially endangered.

Dams and Ecosystems

After dam construction, it is not simple to keep ecosystem parameters upstream and downstream at the same levels as preconstruction conditions. Most frequently, temperature and flow regime are disturbed, particularly if the purpose of the dam is power production, where a reservoir water body is not thermally and hydraulically homogenous. Magnitude and frequency of reservoir fluctuations are rapid and huge. Thermal stratification is quite pronounced. Intake structure position at the dam body is one of the important requirements to reduce the effect of complex processes in a reservoir. Water quality disturbances upstream from a dam are transferred to downstream flow. Enormous daily flow fluctuation and velocity due to hydroelectric power plant operation could have disastrous effects on flora, fauna, and physical properties of a riverbed. To minimize downstream negative environmental impacts, guaranty ecological flow is an essential requirement.

Guaranty ecological flow (or ecological base flow, or environmental flow requirements) is usually one of the controversial requirements in dam construction. Different methods are proposed to define base flow (for instance, the Tennant method). Compared with frequently used sustainable flow or biological minimum flow, the guaranty ecological flow is accepted as the quantity of water flow which guaranties natural ecosystem sustainability [16]. Flora and fauna are the

key parameters to establish balance between all ecological parameters, that is, to preserve ecosystem integrity. Adaptation processes may take place over many years before a new ecological balance is achieved. In the case of deep reservoirs for electricity production, with deep intake structures, the temperature of water downstream from a dam usually is much colder than under natural conditions. Daily, seasonal, and annual flows are quite different than under natural conditions. Consequently, the ecosystem is disturbed, and immediately below the dam structure, the water cannot be used for recreational purposes.

Dams reduce sediment load downstream. If a dam and a reservoir are constructed on a river close to a seacoast, the estuarial effect (saltwater intrusion) upstream is expected, that is, there can be a negative influence of brackish water on native biodiversity. Due to dam construction, the rate of deposition of sediments (fines, sand, and pebble) is reduced. If commercial excavation of sand and pebbles remains as before, the geometry of the riverbed drastically changes. If excavation of sand and pebbles occurred between the dam and the seacoast, the problem becomes more complicated. Excavated sediments cannot be naturally replaced; the river bottom becomes deeper and deeper. As a consequence, the saline water wedge penetrates upstream much faster. This effect can lead to declines in native aquatic vegetation, fish, and amphibian species. Over time, the natural balance is disturbed and brackish water species replace native species. In some cases, saltwater penetration endangers the quality of groundwater near riverbanks creating irrigation and water supply problem.

One of the most serious environmental consequences of dam construction is obstruction to fish migration. Dams are barriers for migratory fishes, such as salmon, trout, sturgeon, alewife, skate, eel, and many others. Construction of fish ways dates from about the seventeenth century in France. Presently, dams are widely equipped with several types of fish ways. These structures are not effective for all fishes or are only partially effective. In the case of downstream migration, mortality of fish passage through power plant turbines or over spillways is significant [17].

Particular problems appear in the case of reversible (pumping) power plants. When the power plant is in pumping regime, fishes can be sucked in by the turbines and transported through the pressure tube and headrace tunnel at the upper compensation reservoir.

Multiple dams along the river considerably worsen the situation for migratory fishes, but in the case of dams in the Glomma River system (Norway) efficient fish ways were constructed at eight dams along 122 km of river [18]. Four dams along the Peconies River (USA) have been equipped with fish ladders. Successful salmon migration has been studied in the case of Snake River dams, Lower Granite dams, and many other locations.

The Iron Gate Dam at the Danube River is not equipped with "fish ways." As a consequence, the caviar productive fish, sturgeon, and skate are unable to migrate from the Black Sea to the Danube River. The 11 dam cascades on the Volga and Kama rivers impede migration north from the Caspian Sea for several sturgeon migratory species including Beluga (Beluga caviar).

Construction of dams can negatively impact wetland ecosystems. Usually, wetlands are extremely rich in diversity of flora and fauna. Reduction of base flow which feeds wetland areas can lead to declines in aquatic vegetation, fish, and birds. Many wetlands are temporary recovery stations for migratory birds. Dam influence may cause reduction of some bird species; however, in some cases dam operation can be easily adapted to support, or even improve, wetland ecosystems.

Dams and Microclimate

As a consequence of damming, estimated total current reservoir surface is more than 400,000 km^2. Largest reservoirs are Lake Volta, Ghana – 8,502 km^2, Aswan Reservoir, Egypt/Sudan – 5,250 km^2, Itaipu Reservoir, Brazil/Paraguay – 1,350 km^2, Ataturk Reservoir, Turkey – 817 km^2, and Keban Reservoir, Turkey – 675 km^2. Due to hot weather or strong winds, evaporation from reservoir surfaces is estimated at about 2 m^3/m^2 per year.

Exact climatological measurements and analyses related to reservoir impact are rare. Measurements taken at the reservoirs at Pourinari and Mornos (Greece) and Bileća Reservoir (BiH) indicate negligible increase of temperature close to the reservoir areas. For more precise conclusions, long-term monitoring of different parameters is necessary.

According to subjective impressions of local people, there is some climatological influence. In karst areas, intensity of deforestation decreased, and fogs are registered much frequently after dam and reservoir construction. Impact of the Krasnojarsk Dam (Russia) on the Yenisei River reaches 200 km downstream and has an influence on local climate by increasing freezing fog.

Induced Subsidence (Collapse)

Origin of subsidence can be natural or induced. Induced subsidence is a consequence of human activities, but mainly due to groundwater extraction, mining, and dam construction [19–22].

Subsidence development is a common process as a consequence of dam construction and reservoir operation. Induced subsidence is a series of spatially independent random events created by reservoir operation. Events such as these are unpredictable and practically instantaneous. Some prominent examples are the following reservoirs: Wolf Creek (USA), Hutovo (BiH), Slano and Vrtac (Montenegro), Tarbela (Pakistan), Mavrovo (FYR Macedonia), Perdica (Greece), Hammam Grouz (Algeria), Kamskaya (Russia), Lar (Iran), Keban (Turkey), North Dike (Florida, USA), and Huoshipo Reservoir (China).

Subsidence is induced by extensive water level fluctuation in reservoirs and results in extensive water leakage. In some cases, subsidence occurred after many years of successful operation. In the case of Hammam Grouz (Fig. 4), subsidence occurred after 17 years, and in the case of Mavrovo, after 25 years.

In some cases, subsidence creates considerable environmental impact. For instance, the Mavrovo reservoir collapse resulted in heavy damage to local roads and surrounding houses.

In the case of Kamskaye Reservoir (Russia), after dam construction the dissolution process in gypsum has intensified in the vicinity of the reservoir [23]. During the period 1956–1961, 11 collapses occurred. Prior to dam construction, in the same area, only two collapses were registered during the previous 50 years.

Spring Submergence Due to Dam Construction

Submergence of large springs by artificial reservoirs and the consequences on environment and reservoir integrity are frequently discussed. After construction of the 185-m high Oymopinar Dam (Turkey), the large Dumanly spring Qmin = 35.6 m^3/s was flooded by 120 m of water head at maximum storage level; the Trebišnjica Spring (BiH), Qav = 80 m^3/s was flooded by 75 m of water column; the Neraidha Spring (Greece), Q = 10 m^3/s, was flooded by 40 m of Poliphiton Reservoir; the 220-m high Piva Dam (Montenegro) the large Pivsko Oko Spring, Qav = 25.5 m^3/s, was flooded by 70 m; the Rama spring zone (BiH) was submerged by 40–60 m of water column; and the Yarg Spring (Iran), Qav = 0.7 m^3/s was flooded after construction of the Salman Farsi Dam (136 m) by 27 m; Oko Spring (BiH) was flooded after construction of 35-m high Gorica Dam by 17 m of water column [24]. The Bel Spring (Iran) will be submerged by construction of the 230-m high Darian Dam.

The most important questions to be answered are as follows: Is the water used for water supply or water bottling, if so, how to keep quantity and quality at acceptable levels?; Are water losses from the reservoir possible?; What is the possible submergence effect on the hydrologic regime in upstream areas?; How could spring submergence affect induced seismicity in surrounding areas?. In many analyzed cases, spring submergence does not increase considerable environmental consequences.

One specific case is Bel Spring in Iran. Spring discharges vary between 150 l/s and 10 m^3/s. Two companies use water from the Bel Spring for bottling. By construction of the Darian Dam, the Bel Spring will be exposed to a water head of more than 150 m. The question is how to protect Bel Spring water for bottling (quantity and quality). A particular tapping structure is designed behind the spring to prevent reservoir water influence and preserve groundwater potential for bottling during reservoir operation.

The Oko Spring is the only water supply source for the town of Trebinje (BiH, 20,000 people). After construction of the Gorica Dam, the tapping structure is above the reservoir level. Three large diameter wells were drilled into the karst channel situated 25 m deeper than reservoir bottom. Impact of reservoir water to quality of potable water occurs only during extremely fast impounding of the reservoir.

Dam Engineering and its Environmental Aspects. Figure 4
Hammam Grouz Dam, Algeria. Subsidence occurred after 17 years of dam operation

Environmental Aspects of Dams in Karst

The complexity of karst presents a great variety of risk for dams and reservoirs in karst. The crucial role of dams and tunnels in karst is dewatering of temporarily flooded karst depressions, for water transfer from one catchment (or political entity) to another and for electricity production. By dam construction in karst regions, some temporarily flooded depressions are changed to permanent reservoirs or, in other cases, to farmland areas. By applying different geotechnical measures, the karstified and pervious riverbeds are transformed into permanent river flows.

Construction of dams and reservoirs in general has considerable influence on regime and quality of surface and underground water downstream. In the case of karst, impacts are sometimes registered at remote springs at distances of 10–30 km, leading to local and transboundary environmental and political problems.

In many cases, the purpose of dam structures is rerouting and transfer of water from one catchment to another. These solutions create conflicts between owners of the dams (reservoirs) and users of the springs. This situation is especially delicate if the reservoir and springs are in different political regions. For example, by construction of Grančarevo and Gorica dams at BiH, the average yearly discharge of Ombla Spring (Croatia) has been reduced from Qav = 33.8 m^3/s to Qav = 24.4 m^3/s. A large part of the water is transferred through the headrace tunnel toward the power plant of Dubrovnik located in Croatia. There was no change in the minimal discharge of Ombla Spring. By construction of dams, the karst aquifer was starved of about 4 billion cubic meters of water annually as a result of rerouting through the tunnels and paved channels for power production and drainage of many swallow holes [24].

Karst underground is very rich with various fauna. Often, as a result of dam construction in karst, a large volume of caverns in the aeration zone are flooded, or temporarily flooded karst channels become permanently dry. In both cases, cave habitat for a number of rare and endemic species is endangered, for example, Normandy Dam (Tennessee, USA), Melond Dam (California, USA), Scrivener Dam (Australia), Grančarevo and Gorica Dams (BiH), and Seymareh Dam (Iran).

A specific example is Popovo Polje in BiH (Fig. 5). Flooding under natural conditions before dam

Dam Engineering and its Environmental Aspects. Figure 5
Very pervious Trebišnjica riverbed covered by shotcrete. At hillside is visible line indicating flood level in natural conditions

construction reached a height of 40 m in the lowest section of the polje; the polje was under water an average of 253 days and was dry only 112 days. During maximum flood, 7,500 ha were under water. During dry periods the Trebišnjica River was dry also because of 65 m^3/s of seepage along the 65 km of riverbed [25]. The only variety of maize that grows here is called "hundred-day maize." In 100 days, it sprouts, grows, and bears fruit. After dam construction and the increase in impermeability of the riverbed, floods were almost eliminated and huge areas of arable land were created. New infrastructure (plantations, roads, irrigation canals, and settlements) changed complete environmental properties of the entire area.

Dewatering of the temporarily flooded Popovo Polje (BiH) has negative influence for aquatic organisms that inhabit temporary underground lakes during the dry period and ephemeral lakes during flood season. An example is the Gaovica fish (*Paraphoxinus ghetaldi*) which spends dry months in numerous siphon lakes and pools of the underground karst. During flood periods, the fish leaves the underground through karst channels and openings of estavelles. For the duration of inundation, the fish lives in the intermittent lakes at surface. For centuries, fishing at openings of estavelles was an important tradition and food source for inhabitants of Popovo Polje. By construction of two dams (Grančarevo and Gorica dams), the water regime has drastically changed; the Gaovica fish lost connections with the surface in most locations and is now threatened with extinction. As a consequence of the same project, the large concentrations of endemic worm *Mariphugia cavatica*, mollusk *Kongeria*, and cave-dwelling aquatic endemic species *Proteus anguinus*, known as "human fish" are seriously endangered [26]. Dam and reservoir construction decreased the activity time and discharge of a series of temporary and submarine springs along the seacoast. The operation of a commercial oyster and mollusk farm has been threatened because of reduced freshwater outflow through the submarine springs.

One of the richest caves with various fauna in the world is in the same area, Vjetrenica cave. Approximately 110 species have been identified in this cave. More than ten species are known only from this cave or the immediate vicinity [27].

From an environmental point of view, underground dams and reservoirs in karst have benefits. The main advantages of these unique structures are as follows: Arable land is not disturbed, water is not thermally stratified, water quality remains high and constant, catastrophic dam failure is not a possible outcome, and the landscape is not disturbed. In the karst regions of China, more than 20 underground reservoirs have been created with storage capacities between 1×10^5 and 1×10^7 m^3. Their purposes are water supply, irrigation, industry, and electricity production [28]. According to Yuan [29], in the Xiashi district (Guizhou Province) 16 underground dams have been constructed for irrigation of 3,624 acres of farmland [29].

To control concentrated infiltration through the large ponors (swallow holes) a specific type of dam is constructed in karst areas – cylindrical dams. The purpose of cylindrical dams is to prevent natural plugging of large ponors, that is, to ensure fast dewatering of floodwater from farmlands. Some cylindrical dams constructed at Peloponnesus (Greece) in fourteenth century are still operational (Fig. 6).

In the Dinaric karst area (BiH), cylindrical dams were used from ancient time until the middle of the twentieth century for mills. These dams were constructed above the large ponors at the riverbanks. Water sinking into the ponor propels wooden turbines and millstones. Mills were equipped with simple intake structures and gates to control quantity of water. Recently, cylindrical dams were used to prevent leakage from reservoirs in the case of large ponors or estavelles in Chinese karst and at some other locations.

During filling of Salanfe Reservoir (52-m high dam, Switzerland), new thermal springs appears in the Val d'Illes valley, at a distance of 8 km (1953). Springs are related to the leakage from Salanfe reservoir. The reservoir has never been filled completely since the first phase of impoundment [30].

Reservoirs in karst may fail to fill despite an extensive investigation program and sealing treatment. Dried reservoirs or reservoirs with unacceptable heavy leakage are common in many karst regions of the world: Hales Bar Dam (USA), Montejaque (Spain), Vrtac (Montenegro), Lar (Iran), May (Turkey), Perdikas (Greece), Wolf Creek (USA), Apa Reservoir (Turkey), and many others. A distinctive example is the

Dam Engineering and its Environmental Aspects. Figure 6
Ancient dam at Peloponnesus, Greece, to protect natural plugging of large ponor (swallow hole)

Dam Engineering and its Environmental Aspects. Figure 7
Montejaque Dam, abandoned due to huge leakage

Montejaque Dam in Spain. The dam was abandoned because of huge leakage from reservoir, and cavernous system downstream from the dam is presently used as a training area for speleologists (Fig. 7).

Further Directions

Optimal strategies for water resources development are a key requirement for socioeconomic development. With increasing demands on energy, particularly in many underdeveloped countries, dams and reservoirs are still necessary structures. The environment has been modified by dam construction with possible detrimental impacts. In most instances, impacts are positive: flood control, irrigation, water supply, power production, infrastructure improvement, reduction of deforestation, recreation, fishing, and many secondary benefits. Some negative impacts cannot be avoided: population replacement, inundation of arable land, historical and cultural monuments, influence on survival of endemic species and migratory fishes, deterioration of aquifers, the possibility of triggered seismicity, collapses, and similar events.

An important issue is how to keep the balance between the necessity for development and preservation of environment. Optimal environmental protection requires a multidisciplinary approach, including close cooperation of a wide spectrum of geologists, civil engineers, biologists, chemists, hydrologists, hydrogeologists, archeologists, sociologists, and many others. The ultimate aim is identification of crucial parameters that define causes and consequences between human activities (dam construction) and resulting impact on environment (cause-and-effect relations). Criteria for determining environmental protection, as well as regulatory procedures are important elements in the process.

Particularly, sensitive and complex is construction of large dams and reservoirs in highly developed karst because the majority of water flows through the underground karst conduits. Dam impacts on environment in karst may be unpredictable, occur rapidly, and may be unique. Similar situations are seldom, if ever, repeated. To expect the unexpected, should be permanently kept in the mind as the basic philosophy of dam construction in karst. The major aims of proper

planning of water resources systems in karst terrain are to minimize negative and to maximize positive environmental impacts by keeping water at surface level as much as possible.

Bibliography

Primary Literature

1. Liu J (1980) Changjiang (Yangtze River) China's largest river. Foreign Language Press, Beijing
2. Iranian National Committee on Large Dams (1993) A general view on Iranian dams, past-present-future. Ministry of Energy, Teheran
3. Tennessee Valley Authority (1949) Tennessee valley authority projects, geology, and foundation treatment. Technical report no 22, Knoxville, Tennessee
4. Flagg CHG (1979) Geological causes of dam incidents. Bull Int Assoc Eng Geol 20:196–201
5. Cooper AH, Calow RC (1998) Avoiding gypsum geohazards: guidance for planning and construction. British Geological Survey. Technical report WC/98/5
6. Gutierrez F, Desir G, Gutierrez M (2003) Causes of the catastrophic failure of an earth dam built on gypsiferous alluvium and dispersive clays, Altorricon, Huesca Province, NE Spain. Environ Geol 43:842–851
7. Schuster RL (2006) Interaction of dams and landslides. Case studies and mitigation. U.S. Geological survey professional paper 1723, p 107
8. Moon A (1997) Predicting low probability rapid landslides at Roxborough Gorge, New Zealand. In: Ctuden D, Fell R (eds) Landslides risk assessment. Balkema, Rotterdam, pp 272–284
9. Selli R, Trevisan L, Carloni GC, Mazzanti R, Ciabatti M (1946) La frana del Vaiont. Annali del museo geologico di Bolona, Bolongna
10. Lu Y (2001) Rational exploitation of resources and prevention of geohazards in karst regions. Acta geologica sinica (english edition). J Geol Soc China 75(3):239–248
11. Chronology of major tailings dam failure, 1960–2011 (prepared on the basis of) Bulletin 121, Published by United Nation Environmental Programme Division of Technology, Industry and Economics and International Commission on Large Dams, Paris 2001, p 144, updated March 2011, (221 tailings dam incidents)
12. Rico M, Benito G, Salgueiro AR, Diez-Herrero A, Pereira HG (2008) Reported tailings dam failures. A review of the European incidents in the worldwide context. J Hazard Mater 152:845–852, Elsevier
13. Bozovic A, Wieland M (2004) Reservoirs and seismicity, state of knowledge. ICOLD Committee on Seismic Aspects of Dam Design. Commission Internationale des Grands Barages, Haussmann, Paris
14. Kiernian K (1988) Human impacts and management responses in the karsts of Tasmania. In: Proceedings of the international Geographical Union, Study Group Man's Impact on Karst, Sydnay
15. Smith D (2000) Protest grow over plan for more Turkish dams. National Geographic News
16. Djordjević B, Dašić T (2007) Ecological guarantied discharge downstream from the hydropower plants. J Elektriprivreda 1. Belgrade, 5–13
17. Larinier M (2001) Environmental issues, dams and fish migration. In: Marmulla G (ed) Dams, fish and fisheries: opportunities, challenges and conflict resolution. FAO fisheries technical paper, 419. Food and Agriculture Organisation of the United Nations, Viale dell Terme di Caracalla, Rome
18. Linlokken A (1993) Efficiency of fishways and impact of dams on the migration of Grayling and Brown Trout in the Glomma river system, South-Eastern Norway. Regul River Res Manag 8:145–153, Wiley
19. Lamoreaux PE, Newton JG (1986) Catastrophic subsidences: an environmental hazard, Shelby County, Alabama. Environ Geol Water Sci 8(1/2):25
20. Hua SL (1987) Pumping subsidences of surface in some karst areas of China. In: Symposium on human influence on karst, Postojna
21. Tolmachev V, Ilyin A, Gantov B, Leonenko M, Khomenko V, Savarenski I (2003) The main results of engineering karstology research conducted in Dzerzinsk, Russia (1952–2002). In: Beck B (ed) Sinkholes and the engineering and environmental impacts of karst. Geotechnical special publication no 122. Alexander Bell Drive, Reston, Vurginia
22. Gutierrez F (2010) Hazards associated with karst. In: Alcantra-Ayala I, Goudie A (eds) Geomorphological hazards and disaster prevention. Cambridge University Press, Cambridge
23. Maximovich NG (2006) Safety of dams on soluble rock (The Kama hydroelectric power station as an example). On Russian. The Russian Federal Agency for Science and Innovations, Perm
24. Milanović P (2004) Water resources engineering in karst. CRC Press, Boca Raton
25. Milanović P (1981) Karst hydrogeology. Water Resources, Colorado
26. Milanović P (2002) The environmental impacts of human activities and engineering constructions in karst regions. Epizodes J Int Geosci 25(1):13–21
27. Sket B (2003) Vjetrenica Cave. In: Lučić I, Sket B (in Croatian), Zagreb-Ravno
28. Lu Y (1986) Some problems of subsurface reservoirs constructed in karst regions of China. Institute of Hydrogeology and Engineering Geology, Beijing
29. Yuan D (1990) The construction of underground dams on subterranean streams in South China karst. Institute of Karst Geology, Guilin, p 62
30. Roth P (1994) Reservoir induced earthquakes and thermal springs in valais. Proseis AG, Zurich

Books and Reviews

Abraham S (2008) Determining ecological baseflow needs for stream and springs in arid and semi-arid regions. Geological Society of America, Abstracts with programs 40(6): 87

Arthur HG (1977) Teton dam failure. In: The evaluation of dam safety: engineering foundation conference proceedings, ASCE, New York, pp 61–71

Avakyan AB, Romashkov EG, Chestnaya ET (1979) Fishways and dams (Trans: Gidrotechnickoe Stroitelstvo, 1970). Power Techn Eng 4(11)

Bergkamp G, McCartney M, Dugan P, Mcneely J, Acreman M (2000) Dams, ecosystem functions and envoronmental restoration: thematic review II. I World Commission on Dams, Cape Town

Bizer JR (2000) Avoiding, minimizing, mitigating and compensating the environmental impacts of large dams. IUCN, Gland

Bozovic A (1974) Review and appraisal of case histories related to seismic effects of reservoir impounding. Eng Geol 8(1–2):9–27

Chandler RJ, Tosatti G (1995) The Stava talings dam failure, Italy, July 1985. Proc Inst Civ Eng 113:67–79

Cogels FH, Coly A, Niang A (1977) Impact of dam construction on the hydrological regime and qualirty of a Sahelian Lake in the River Senegal Basin. Regul River Res Manag 13(1):27–41

Collier M, Webb RH, Schmidt JC (1996) Dams and rivers: a primer on the downstream effects of dams. US Geological survey circular 1126. US Geological Survey, Tuscon

Djordjevic B (1990) Cybernetics in water resources management. WRP, Colorado

Dobry R, Alvarez L (1967) Seismic failures of Chilean tailings dams. J Soil Mec Found Div 93:237–260

Fell R, Bowles DS, Anderson LR, Bell G (2010) The status of methods for estimation of the probability of failure of dams for use in quantitative risk assessment. In: International commission on large dams, 20th congress, Beijing

Ford D, Williams P (2007) Karst hydrogeology and geomorphology. John Wiley & Sons, Chichester

Fourie AB, Papageorfiou G, Blight GE (2000) Static liquefaction as an explanation for two catastrophic tailings dam failures in South Africa. In: Tailings and mine waste 2000, proceedings of the seventh international conference on tailings and mine waste 2000, Fort Collins, 23–26 Jan 2000, Balkema, Rotterdam, pp 149–158

Gupta HK (1992) Reservoir-induced earthquakes. Elsevier, Amsterdam

Helms SW (1981) Jawa, lost city of the black desert. Methuen and Co Ltd, USA

Holmgren R (2000) Experiences from the tailing dam failure at the boliden Atik and legal consequences, international report, County Administration of Norrbotten, Sweden

ICOLD Bulletin 99 (1995) Dam failure, statistical analysis. Commission Internatrionale des Grandes Barrages – 151, Paris

Jackson CD, Marmulla G (2001) The influence of dams on river fisheries. In: Marmulla G (ed) Dams, fish and fisheries: opportunities, challenges and conflict resolution. FAO, Fisheries technical paper, 419. Viale dell Terme di Carcalla, Roma

Jobin WR (1999) Dams and disease: ecological design and health impacts of large dams, canals, and irrigation systems. Taylor & Francis, London. ISBN 0419223606 (2)

Keffer ML, Bjornn TC, Peery CA, Tolotti KR, Ringe RR, Stuehrenberg L (2003) Adult spring and summer Chinook salmon passage through fishways and transition pools at Bonneville, McNary, Ice. A report for project MPE-P-95-1. U.S. Geological Survey, Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho, Moscow, Idaho 83:844–1141

Lafitte R (2001) Ethics and dam engineers. Int J Hydropower Dams 4:58–59

Lamoreaux PE (1989) Water development for phosphate mining in a karst setting in Florida – a complex environmental problem. Environ Geol Water Sci 14(2):117–153

Liu JK, Yu ZT (1992) Water quality changes and efects on fish population in the Hanjiang River, China, following hydroelectric dam construction. Regul River Res Manag 7(4): 359–368

McCartney M (2009) Living with dams: managing the environmental impacts. Water Policy 11(Supp 1):121–139, IWA Publishing

McDonald MJ, Muldowny J (1982) TVA and the dispossessed: the resettlement of population in the Norris Dam. University of Tennessee Press, Knoxville

McKartney MP (2007) Decision support systems for large dam planning and operation in Africa. Working paper 119. International Water Management Institute, Colombo

Nations Environmental Programme (UNDP) Division of Technology, Industry and Economics (DTIE) and International Commission on Large Dams (ICOLD) (2001) Risk of dangerous occurrences, lessons learnt from practical experiences, Bulletin 121. Nations Environmental Programme (UNDP) Division of Technology, Industry and Economics (DTIE) and International Commission on Large Dams (ICOLD), Paris, p 144

Odum EP (1969) The strategy of ecosystem development. Science 164:262–270

Perman RC, Packer D, Coppersmith KJ, Kneupfer PL (1983) Collection of data bank on reservoir-induced seismicity. USGS contract no 14-08-0001-19132

Petrović P (2010) "The Kosheish Dam" has it really existed or not? Vodoprivreda no 246-248, Belgrade

Shen CK, Chen HC, Chang CH, Huang LS, Li TC, Yang CY, Wang TC, Lo HH (1973) Earthquakes induced by reservoir impounding and their effects on Hsingfeng dam. Sci Sin 17(2):232–272, Beijing, China

Simpson DW (1988) Two types of reservoir induced seismicity. Bull Seismol Soc Am 78:2025–2040

Tanaka E (2007) Protecting one of the best roman mosaic collections in the world: ownership and protection in the case of the roman mosaics from Zenguma Turkey. Stanford J Archeol 5:183–202

United Nations Environment Programe (1996) Environmental and safety incidents concerning tailings dams at mines: results of a survey for the years 1980–1996 by mining journal research services. United Nations Environment Programe, Industry and Environment, Paris, p 129

USCOLD (1997) Reservoir triggered seismicity. The sixtinth USCOLD Anual Meeting, LA. US Society of dams. Debver, CO

U.S. Army Corps of Engineers (1996) Harbour, and lower granite dams. Technical report 2003-5. U.S. Army Corps of Engineers, Portaland and Walla Walla District

U.S. Committee on Large Dams – USCOLD (1994) Tailings dam incidents. U.S. Committee on Large Dams – USCOLD, Denver, 82 p. 1-884575-03-X

Van Niekerk HJ, Viljoen MJ (2005) Causes and consequences of the Merreispruit and other tailings-dam failures. Land Degrad Dev 16:201–212

Vick SG (1990) Planning, design, and analysis of tailings dams. Bitech, Vancouver

Wieland M Dam safety aspects of reservoir-triggered seismicity. In: Wieland M, Ren Q, Tan SY, Taylor, Francis (eds) Book New Developments in Dam Engineering. pp 95–100

Daylight, Indoor Illumination, and Human Behavior

JOHN MARDALJEVIC
Institute of Energy and Sustainable Development, De Montfort University, Leicester, UK

Article Outline

Glossary
Definition of the Subject
Introduction
Buildings and Daylight
Guidelines, Measures, and the Evaluation of Daylighting
Human Factors and Daylight
Recent Advances in Daylighting
Future Directions
Bibliography

Glossary

BTDF The bidirectional transmission distribution function. The BTDF characterizes the distribution in luminous output across the full hemisphere of transmitted rays and usually needs to be determined for every incident direction. It is needed to simulate the performance of complex glazing materials and systems.

Climate-based daylight modeling The prediction of various radiant or luminous quantities (e.g., irradiance, illuminance, radiance, and luminance) using sun and sky conditions that are derived from standardized annual meteorological datasets.

Daylight The totality of visible radiation originating from both the sun and the sky.

Daylight factor The ratio of internal illuminance to unobstructed external illuminance under the CIE standard overcast sky.

Daylight metric Some mathematical combination of (potentially disparate) measurements and/or dimensions and/or conditions of daylight represented on a continuous scale.

Illuminance The total luminous flux incident on a surface per unit area. It is a measure of the intensity of the incident light, wavelength-weighted by the eye's sensitivity to correlate with human brightness perception. SI unit: lux or lumens per square meter.

Luminance Photometric measure of the luminous intensity per unit area of light traveling in a given direction. It describes the amount of light that passes through or is emitted from a particular area, and falls within a given solid angle. The SI unit for luminance is candela per square meter.

Definition of the Subject

Daylight in buildings is the natural illumination experienced by the occupants of any man-made construction with openings to the outside, e.g., dwelling and workplace. The quantity and quality of daylight in buildings is continually varying due to the natural changes in sun and sky conditions from one moment to the next. These changes have components that are random (e.g., individual cloud formations), daily (i.e., progression from day to night), and seasonal (e.g., changing day length and prevailing weather patterns). For any given sky and sun condition the quantity and character of daylight in a space will depend on the size, orientation, and nature of the building apertures; the shape and aspect of the building and its surroundings; and the optical (i.e., reflective and transmissive) properties of all the surfaces comprising the building and its surroundings.

The purpose of the very earliest shelters – the forerunners of buildings – was to protect from the elements. The first buildings to include deliberate elements of daylighting design were often places of worship, many of which survive to this day. Only when glass became relatively commonplace in the

seventeenth century did the provision of daylight for everyday buildings become a consideration. Window design was invariably tailored to the prevailing climate, e.g., small with deep reveals for locales where the solar component of daylight needed to be controlled to prevent overheating. As the cities of the industrialized world became more populous, building densities increased and the provision of daylight for buildings became a planning issue. This eventually resulted in the formulation of the daylight factor which was intended to be a measure of the daylighting potential of a building, and which could be predicted at the design stage using a variety of methods. Devised in the first half of the twentieth century, the daylight factor still forms the basis of many guidelines and recommendations for building design, notwithstanding the fact that a daylight factor value is, by definition, completely insensitive to the building orientation and any consideration of climate.

Advances in glass making and window technology allowed architects to design buildings where the perimeter wall could be almost entirely glazed. Commercial buildings in particular became larger with deeper plan designs so that, despite the highly transparent facade, many occupants were situated far from the windows and so received little daylight. As conspicuous icons symbolizing modernity and prosperity, these designs became the exemplars for architects all over the world, and now many cities feature highly glazed buildings regardless of the local climatic conditions. Thus the daylighting characteristics of office buildings in particular tended to be dictated by considerations of architectural style rather than climate-adapted design. These trends were not hindered by the continued reliance on the daylight factor as an evaluative scheme since the measure is itself climate and orientation insensitive.

For the majority of buildings, it is incumbent on the occupants to moderate the internal daylight conditions using some form of blinds or shades. Occupants will deploy blinds/shades in an effort to moderate the internal environment according to their perceptions of both visual discomfort (e.g., daylight glare) and thermal discomfort (e.g., to avoid direct sun) which vary greatly from person to person. Also, once deployed, blinds/shades will tend to remain closed long after the external condition has passed. Thus it is common to see blinds closed for much of the occupied time. Consequently, the potential to exploit daylighting is often not realized because the blinds are left closed most of the time and the electric lights are left switched on.

Toward the end of 1990s, the daylighting of buildings began to achieve greater attention for a number of reasons. The two most important drivers were:

1. The widespread belief that the potential to save energy through effective daylighting was greatly underexploited.
2. The emergence of data suggesting that daylight exposure has many positive productivity, health, and well-being outcomes for building occupants.

The first of these concerns originated in the 1970s following the energy crisis and culminated with the widely accepted need to reduce carbon emissions from buildings in order to minimize the anticipated degree of anthropogenic climate change. This in turn led in the 1990s to the formulation of guides and recommendations to encourage the design and construction of low-energy buildings and also for the retrofit of existing buildings. All these guides contain recommendations on daylighting, invariably founded on the daylight factor or an equally simplistic schema such as glazing factors. The productivity, health, and well-being effects related to daylight exposure are not yet fully understood, and it is not yet known what the preferred exposure levels should be nor if existing guidelines would be effective for these quantities.

Almost concurrent with the emergence of the two key drivers noted above were a major advancement in the way daylight in buildings could be modeled and the development of numerous new glazing systems and materials to better exploit daylighting in buildings. These developments are expected to lead to significant changes in the way that daylight in buildings is both evaluated (through modeling) and exploited (by new glazing systems and materials).

Introduction

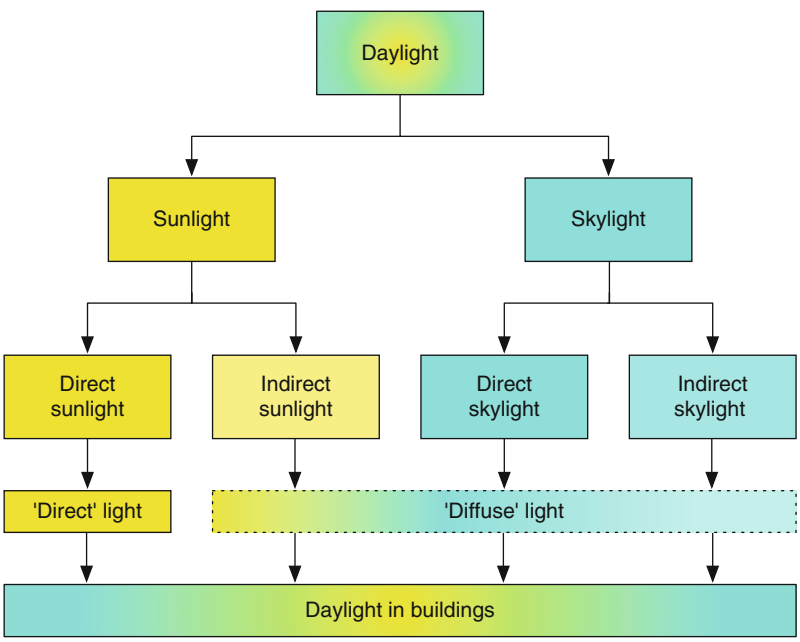
Daylight is generally taken to be the totality of visible radiation originating from the sky and, when visible, the sun during the hours of daytime. The source of all daylight is in fact the sun. Scattering of sunlight in the atmosphere by air, water droplets, and dust gives the sky the appearance of a self-luminous hemispherical

source of light. Sunlight is commonly referred to as direct light since it appears to originate from a small source and can be highly luminous casting sharp shadows. The sky, however, is an extended source of illumination that casts only soft shadows and so skylight is commonly referred to as diffuse light. In building science, diffuse light includes also light from the sky and/or the sun that has undergone one or more reflections from surfaces that are generally not mirrorlike in character.

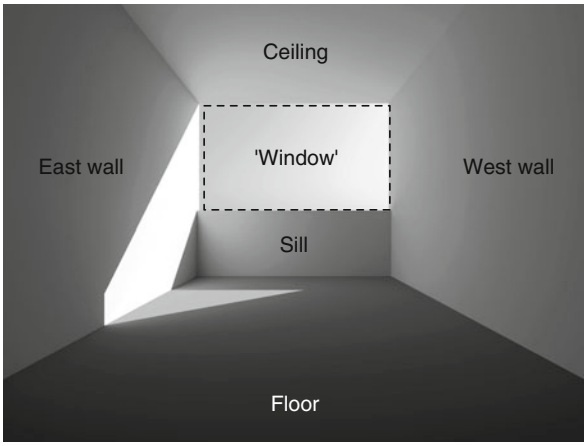
Daylight may arrive at a point inside a building either directly or indirectly from the luminous source, i.e., from the sun or from the sky. Direct illumination generally results from having an unobstructed view of the source. Indirect illumination is when the light arrives at the point following one or more reflections. Thus, strictly speaking, there are direct and indirect components of illumination from both the sun and the sky (Fig. 1). Although the sun and the sky are both luminous sources, direct sunlight when present is given special consideration because of the small angular size of the sun and its potentially large contribution to illumination (and also its heating effect). Thus illumination from direct sunlight is commonly referred to as "direct light." In contrast, light from the sky – arriving either directly or indirectly – is commonly referred to as "diffuse light." Sunlight that has undergone one or more diffuse reflections is also commonly referred to as "diffuse light." Note that the mode of reflection of the direct sunlight is important: a specular (or "mirror") reflection of sunlight will produce a *redirected* beam of direct light rather than diffuse light. For reflections (and transmissions) that are part-specular and part-diffuse, the distinction between direct and diffuse light can become lost. Reflections can occur either internal or external to the building space under consideration.

Components of Daylight Illumination Indoors

The image shown in Fig. 2 is a simulation of the daylight distribution in a simple space under clear sky conditions. The viewpoint of the virtual camera is "looking" toward the south-facing window. The external conditions are those of a clear sky with sun, with the sun 30° west (i.e., to the right) of the view direction, and at an altitude of 45°. The direct sun illumination on the east wall and the floor is clearly visible. The space

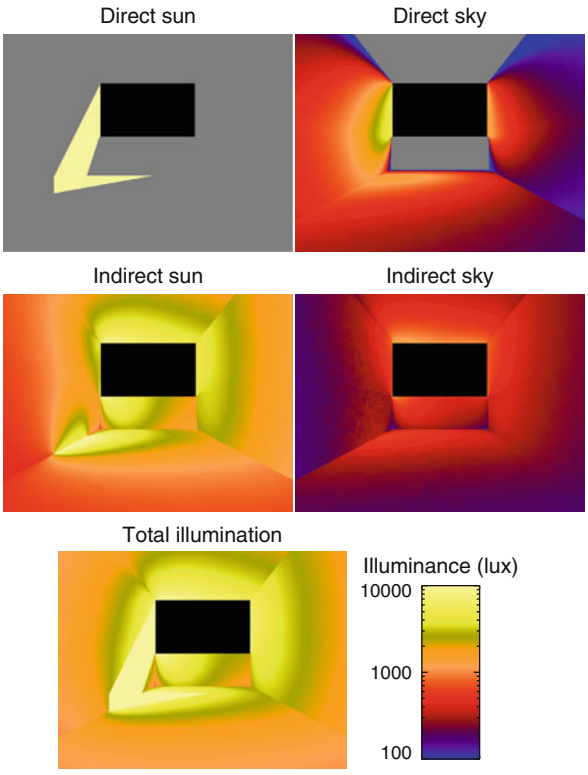


Daylight, Indoor Illumination, and Human Behavior. Figure 1
Components of daylight illumination in buildings



Daylight, Indoor Illumination, and Human Behavior. Figure 2
Simulated image of a simple space illuminated by sunny clear sky conditions

is 3.0 m wide, 9 m deep and 2.7 m high and it "sits" on a ground plane that is 100 m^2. The reflectance values assigned to the surfaces are typical of those commonly used for office buildings and the external ground plane. No glazing was used in the window aperture to eliminate potentially distracting reflection patterns for the example that follows. The simulation of the space was actually carried out in stages so that the individual components of illumination described above could be computed and shown individually. The image in Fig. 2 shows how the space might appear to the eye, in other words it is a representation of the predicted surface brightness (or luminance) of the scene. The light that is incident on a surface is termed the illuminance and is a measure of the illumination (i.e., the light arriving) at a particular point – this is what we would measure with a light or lux meter. The images in Fig. 3 show the distribution and relative proportions of the four components of daylight as illuminance values, together with a fifth image that shows the total illuminance (i.e., the sum of the four components). The magnitude of the illuminance is shown using color and the relation between the two can be read from the legend. Note that a logarithmic scale is used that covers the range from 100 to 10,000 lux. Areas where the illuminance value was zero are shaded gray and the window aperture has been shaded black.



Daylight, Indoor Illumination, and Human Behavior. Figure 3
Components of illumination for the simple space shown in Fig. 2

Direct Components The direct sun image shows the pattern of sun illumination on the East wall and the floor. The illuminance value on the wall is around 40,000 lux (values higher than the legend maximum of 10,000 lux will all have the same shade). Because the scene contains no obstructions, all of the surfaces visible in the image except the ceiling and the sill (Fig. 2) have a direct "view" of the sky and so they receive some illumination directly from the sky. The amount of illumination received at a point depends on:

- The angular size of the patch of sky "seen" through the window (from that point)
- The brightness (i.e., luminance) of the patch of sky
- The transmission properties of the glazing (not used for this illustration)

The areas of wall nearest to the window have the best view of the sky, and they receive the largest

illumination. Note that the illuminance is greater on the East wall because this side "sees" the circumsolar region, i.e., the locus for the sun position and therefore the brightest part of the hemispherical clear sky pattern. The illuminance toward the back of the space (away from the window) reduces gradually as the apparent size of the window diminishes. The illuminance at the highest part of the walls drops off rapidly (i.e., blue shade) as the view of the sky reduces to zero.

Indirect Components The indirect components of illumination show the illuminance that results from multiple light reflections. These reflections usually introduce light from the external environment and also redistribute light within the space. The degree to which this occurs depends on the quantity of incident light (i.e., the direct components) and the reflective properties of the various surfaces. The images showing the indirect components do not include the direct component. The pattern of indirect sun illumination in Fig. 3 is quite complex, but it can be understood with reference to the direct sun pattern. The illuminance resulting from the first reflection of light will generally have the greatest contribution to the indirect component since the intensity of light diminishes with each subsequent reflection. Thus those areas of floor, wall, and ceiling that have the best "view" of the directly illuminated surfaces (inside and out) will have the highest indirect sun illuminance. The west wall "sees" the brightly illuminated sun patch on the east wall (together with the smaller sun patch on the floor) and so it receives a marked quantity of indirect (sun) illumination. For any given size of directly illuminated sun patch, the indirect illumination effect from one on the floor will be less than from one on the wall because the floor has a lower reflectivity. The ceiling will receive indirect sun light from both the sun patches inside the space and the sun-illuminated ground outside. Subsequent reflections will serve to redistribute this, but the bulk of the indirect sun illumination will be at the window end of the space, as is evident in Fig. 3. The indirect component of illuminance from the bulk of the indirect sun illumination will be at the window end of the space, as is evident in Fig. 3. The indirect component of illuminance from the sky shows a gradual diminution going away from the window and toward the back of the space. The slight asymmetry in illumination between the walls can be explained because, at the first reflection, the west wall "sees" the brighter east wall.

This illustration shows how complex the quantity, quality (i.e., direct, diffuse, etc.), and the patterns of natural illumination can be for even the simplest of indoor spaces under naturally occurring daylight conditions. For more realistic spaces, the patterns of illumination will be more complex and, of course, continually changing throughout the day/year.

Buildings and Daylight

The earliest windows were simply holes in the wall or roof of a building. They were providers of daylight and also ventilation, but could be covered with animal hide, cloth, or wood to protect the building occupants from the elements. Daylight would almost always be desirable inside the building, whereas ingress of cold air and rain would generally be avoided. The need for a light transmitting medium that would protect from the elements was first met by translucent materials such as flattened hides and thinly sliced sheets of marble. It was with the invention of glass that the story of daylighting design for buildings truly began.

The first recorded use of glass for windows was not until approximately 100 AD by the Romans. The largest panes that could be manufactured were fairly small. The use of vertical and horizontal dividers – called, respectively, mullions and transoms – increased the size of areas that could be glazed since small pieces of glass could be combined to create large windows. Substantial dividers could additionally form part of the load-bearing structure.

The Industrial Era: 1700s to the Present Day

Glass windows became common in homes in the most developed parts of Europe only in the early seventeenth century. With the advent of improved production techniques in the following century, the cost of glass became less of a limiting factor in its use, though its relative cost compared to other building components was still fairly high. Notwithstanding the high cost of glass, the real cost of artificial light (i.e., as a proportion of the overall household expenditure) was several thousand times what it is today on a per lumen of light basis [1].

The majority of building spaces were side-lit by vertical windows, though the size and arrangement of the windows depended on a number of factors, e.g., intended use, aspect and grandeur or otherwise of the property. Whatever the function or style of the building, window design was, with varying degrees, adapted to the local climatic conditions. Regardless of the size of the window, there would always be a trade-off between daylight provision, heat loss, and solar gain. Internal shutters, often in conjunction with curtains, tended to be used in colder climates where buildings were heated much of the year and heat loss through the windows was significant. In hotter climates where heating was rarely needed, overheating due to undue solar ingress was common in the summer months. Solar gain was moderated by having external shutters which could be closed to cover most or all of the window. Shutters usually had fixed or variable slats that were principally to allow ventilation but also some diffuse light, Fig. 4. Global warming has led to considerations that southern European architectural features such as external shutters may become commonplace in more northerly latitudes in the next 20–50 years.

Rooflights Rooflights began to appear in the mid 1700s when advances in the manufacturing process allowed the fabrication of large sheets of glass at a relatively low cost. Rooflights became a common feature in ordinary houses in the late nineteenth century when mass-produced rooflights with cast iron frames became available. By the late nineteenth century, glass had become relatively inexpensive and was available in good-quality large sheets. Factory buildings since the late 1800s were often designed to provide high levels of natural illumination evenly distributed across the workspace – a goal which is relatively easily achieved in single-story top-lit buildings. The potential for daylighting modern large-span buildings (e.g., storage facilities) was generally underexploited, but energy concerns have led to a rediscovery of this resource.

In domestic buildings, rooflights are now more common and more popular than ever before, particularly in attic conversions and in refurbished historic buildings. In newly designed dwellings, the rooflight is often seen as an integral feature of the daylighting design, and automated opening and blinds operation allows for high-level placement that would have been

Daylight, Indoor Illumination, and Human Behavior. Figure 4
Solar protection using moveable shutters (Dordorgne region, France)

Daylight, Indoor Illumination, and Human Behavior. Figure 5
Rooflights in a modern dwelling (Photo courtesy VELUX A/S)

impractical only 10 years ago, Fig. 5. Depending on the function and occupancy of the space, domestic rooflights can result in significant lighting energy savings [2].

Daylighting in Special-Purpose Buildings

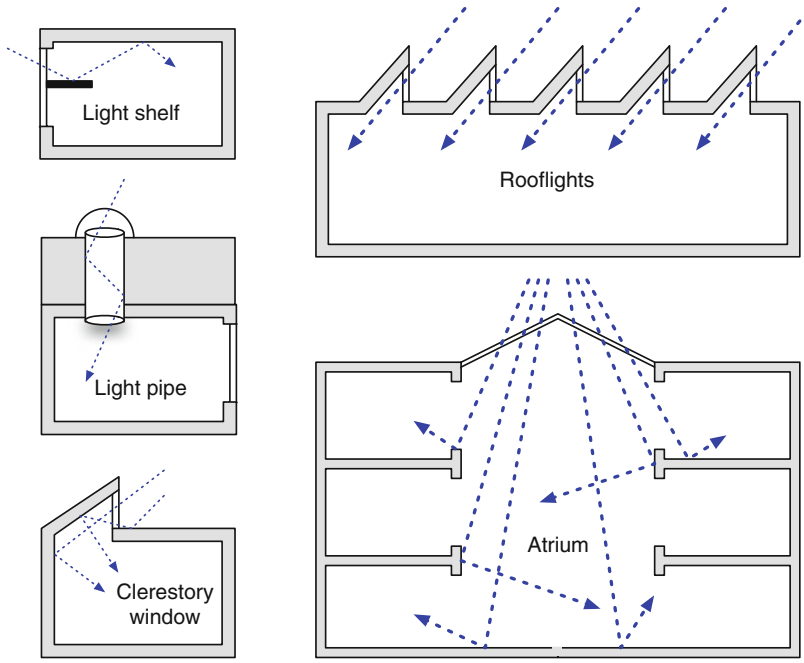
Buildings such as art galleries and museums typically have special requirements for daylighting. Current display practice for highly light-sensitive objects is usually to exclude daylight and use subdued levels of electric lighting commensurate with recommendations for total annual exposure to illumination published by professional bodies and, in some countries, included in legislation. In the wider historical context, this is a relatively recent development since before the 1960s the use and control of light was largely an issue of household practice and museums were usually designed to maximize their use of available daylight through large areas of glazing. From the 1960s onward, daylight has been largely excluded from museum galleries with the exception of those used for art collections (particularly oil paintings) and historic interiors, where its exclusion was felt deleterious to the interpretation of the ensemble of interior and contents. While this move has been partially reversed, with designers expending a great deal of ingenuity trying to blend daylight with new designs of displays and galleries, the success of their work has been variable. In part this is related to our poor understanding of daylight's quantitative performance and the extremely limited techniques, most of them unchanged over half a century, by which daylight in interiors is predicted (see section "The Daylight Factor"). Daylighting design for museums remains one of the more challenging areas of architectural endeavor.

Daylighting Strategies

A daylighting strategy is any building feature that is intended to increase, enhance, or moderate the daylight entering a building. The term is not generally used to describe the typical vertical window with ordinary glazing, though readers may see it used in that way. The strategies can be loosely divided into those that:

1. Employ orthodox construction methods but the building apertures and windows use configurations other than the typical vertical perimeter arrangement
2. Make use of recent and emerging technologies to moderate and/or control the daylight

The first of these, what might be called basic daylighting strategies, are described in this section. The others are described in a later section "Advanced Glazing Systems and Materials". The basic daylighting strategies are essentially building configurations that depart from the norm of vertical glazing; however, they also include, e.g., simple exaggerations of standard building features such as the window reveal to improve the self-shading properties of the window. The various strategies do not occupy distinct categories and many designs feature elements of two or more. There are



Daylight, Indoor Illumination, and Human Behavior. Figure 6
Daylighting strategies

a number of basic daylighting strategies that are commonly used in buildings. A few of the more common ones are illustrated in Fig. 6.

A light-shelf is any horizontal overhang which abuts and divides a vertical window. The upper and lower surfaces of the shelf are usually highly reflective, e.g., painted bright white, though the upper surface may have a mirror finish. The purpose of the light-shelf may be twofold: (a) to redirect daylight deeper into the space through multiple reflections and (b) to offer partial shading from direct sun for those occupants close to the light-shelf.

The clerestory is any window above eye-level height that allows light to penetrate deep into a space. Strictly speaking, the clerestory window in modern buildings is a distinct window aperture, so the upper part of a tall window separated by frame bars would not normally be described as a clerestory window. The effectiveness of a clerestory window for deep penetration of daylight generally improves with increased height of the window. Clerestory windows can feature on the perimeter facade, often above standard height glazing, or further back in the space to better illuminate those areas farthest from the main windows (Fig. 6).

A light-well is a shaft within a building that is open to the outside at the top to admit daylight. The sides of the light-well usually have a bright finish to encourage reflection of light deep into the well. Windows open onto the shaft to admit daylight to spaces that otherwise do not have any direct access to daylight. The level of daylight that a light-well provides to adjacent spaces tends to be fairly small and decreases rapidly away from the opening. The light-well may also serve as a means to encourage natural ventilation for those spaces deep in the core of the building. Where the primary function is ventilation, the structure is usually referred to as an air-shaft. Light-wells and air-shafts were a common feature in early-twentieth-century tenement buildings where the high density placed restrictions on the availability of natural light and air to core spaces.

The light-pipe, also known as light-tube or more generally as a tubular daylight guidance system (TDGS), is an evolution of the light-well design where a conduit transports light through multiple reflections (Fig. 6). A light-pipe is usually comprised of the collector at one end and the diffuser at the other, Fig. 7. The inside surface of the pipe usually has a metallic/mirror finish with a high reflectivity. Light-pipes can be

Daylight, Indoor Illumination, and Human Behavior. Figure 7
Typical light-pipe diffuser designed to appear like a luminaire. Note that the nearby luminaire is daylight sensing and switched off due to the daylight illumination provided by the light-pipe

made to almost any dimensions, though most are circular in cross section and less than 1 m in diameter, and they can be either straight or have minor bends. Light loss in bends will be higher than for the equivalent length of straight pipe, and so are avoided wherever possible. The efficiency of a light-pipe is the ratio of the lumens delivered into the room by the diffuser to the lumens received by the collector, expressed as a percentage [3]. The efficiency decreases with length of the tube, and so light-pipes are rarely longer than 5 m in length. In office buildings, the use of light-pipes is usually restricted to the topmost story, though they may occasionally be used to convey daylight to the next floor below also. For domestic dwellings, they are most commonly used to supply daylight to central areas such as upper-story stairwells that do not receive any light from perimeter windows. Thus the length of the pipe is governed by the distance traversed through the attic space between the ceiling diffuser and the collector on the roof.

Opened in 1993, the Queens Building at De Montfort University (Leicester, UK) is an award-winning low-energy design that incorporates a number of daylighting strategies. The core of the building is illuminated by several rooflights (which double as ventilation openings) and also "borrowed" light from well-daylit adjacent spaces, Fig. 8. The vertical facades on an enclosed courtyard area between two projecting wings have a high reflectivity finish to enhance the inter-reflected component of light that enters the windows, thus ameliorating to a degree the shading effect of the opposing obstructions, Fig. 9. Daylight penetration is encouraged by the use of high-level windows in combination with high ceilings, and a degree of solar protection is provided by having tall narrow windows set in deep reveals, Fig. 10.

The Decline of Climate-Adapted Building Design

Prior to the 1900s, buildings generally incorporated features that evolved from the need to temper the internal conditions in response to the prevailing climate for that locale. In hot climates with a high propensity for sunshine, buildings would be designed to include elements of solar control either by passive or active means. For example, on sun-exposed facades there would be small windows set in deep reveals to provide self-shading (passive control) or perhaps larger

Daylight, Indoor Illumination, and Human Behavior. Figure 8
Rooflights and glazing apertures to allow for light "spillage" from adjacent spaces

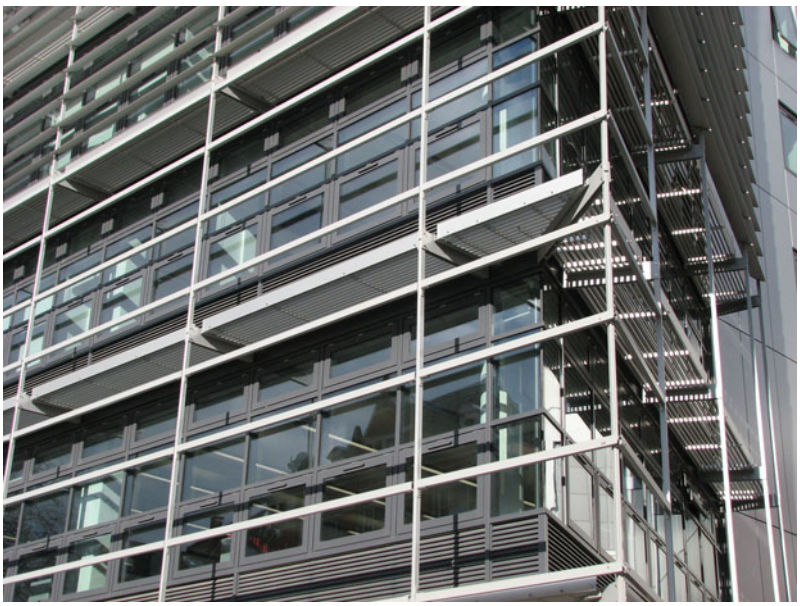
Daylight, Indoor Illumination, and Human Behavior. Figure 9
Highly reflective courtyard

windows with moveable shutters (active control). In less hot/sunny climates, window apertures would tend to be larger and solar control less of an issue – though there would be other concerns such as heat loss. Thus, all buildings contained to varying degrees features in their design that were climate-adapted, and which, over time, became an intrinsic part of that locale's vernacular architecture.

Daylight, Indoor Illumination, and Human Behavior. Figure 10
High-level glazing and deep window reveals

The latter half of the twentieth century witnessed a globalization in architectural form. The highly glazed office tower became the most conspicuous symbol of industrial progress, and hence an inspiration for architects and designers worldwide – regardless of their local climatic conditions. Designers typically rationalized the use of large glazing areas in terms of their daylight provision, but often it was more likely the pursuit of style. Aside perhaps from the possibility of a view to the outside, the daylighting benefit to the majority of occupants in a deep plan space was not great since very few would be close enough to the perimeter to gain any direct benefit in terms of daylight provision. Furthermore, the daylighting potential of highly glazed buildings was often not realized because the manually operated shades/blinds needed to control direct sun were typically left closed long after the external condition had changed. This is the case in fairly temperate climates such as the UK and Northern Europe, but the situation is worse still in sunnier locals. While internal shades do provide occupants with immediate protection from direct sun, they generally have only a limited overall heat rejection effect. Thus, solar radiation, once it has passed through glazing, will inject heat energy to the space which may need to be removed by active cooling (e.g., air-conditioning) if it causes overheating. The almost completely glazed building shown in Fig. 11 exemplifies the total abandonment of climate-adapted design in favor of architectural style. Solar gain will occur across the entire glazed area, and, since the internal shades are dark, virtually all of that solar gain will be instantly reprocessed into heat energy. Thus it is likely that the building will require air-conditioning throughout summer and also at spring/autumn times of the year whenever there is sun. The non-shaded glazed area (i.e., that which could provide daylight but not permanently covered by blinds) occupies approximately one fifth of the total glazed area. Compare that modern design, a tourist office in the Dordogne region, with the climate-adapted traditional design from the same locale shown in Fig. 4. A common practice with highly glazed buildings is to shield the exposed facades with a permanent structure known as a "brise-soleil" (from the French meaning "sun breaker"), Fig. 12. A brise-soleil can be almost any design providing that it offers some degree of solar protection. However, it is true that typical usage of brise-soleil is also more the product of architectural style than resulting from precise performance evaluation of their shading efficiency.

Daylight, Indoor Illumination, and Human Behavior. Figure 11
An almost completely glazed building (tourist office) in the Dordorgne region, France

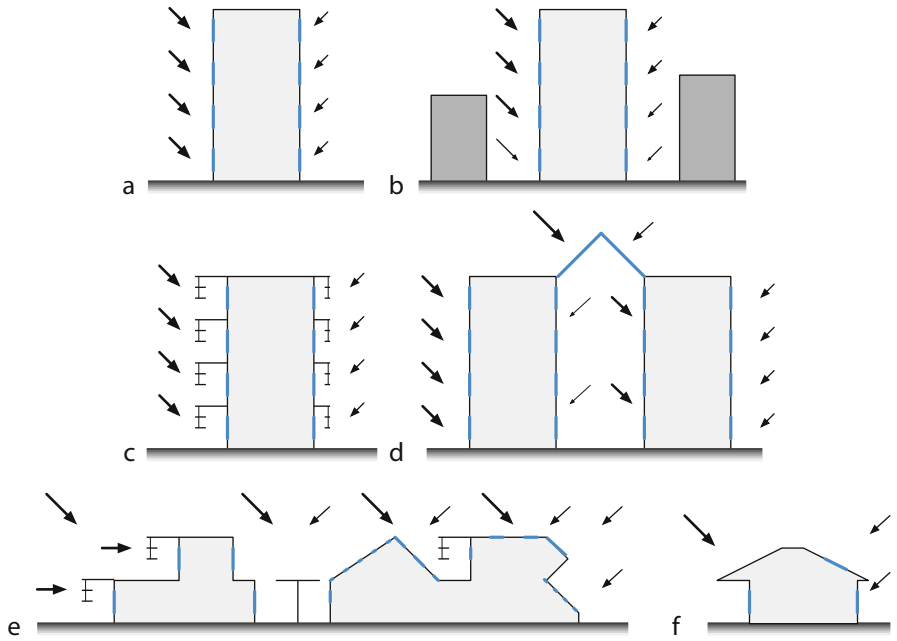


Daylight, Indoor Illumination, and Human Behavior. Figure 12
Solar protection provided by a brise-soleil, Hugh Aston Building, Leicester, UK

The "Well-Tempered" Daylit Environment

A "well-tempered" daylit environment is one where the fixed (i.e., static) architectural form provides both good daylighting and effective solar protection. Thus minimizing – though in practice rarely eliminating – the need for occupants to operate blinds/shades.

The potential for the fixed form to temper the daylighting of the space depends on the building type,



Daylight, Indoor Illumination, and Human Behavior. Figure 13
The "well-tempered" daylit environment

specifically on the richness/variety of the architectural form. This is illustrated by the schematic showing various building types given in Fig. 13. The weight of the arrows is used to indicate the degree of exposure to daylight, i.e., from both the sun and the sky. For the typical rectangular office building (in the northern hemisphere) the south-facing facade will be exposed to direct sun (Fig. 13a). For this type of building the scope to temper the daylit environment is limited to the manipulation of a few basic building parameters, e.g., glazing ratio, transmissivity, etc. Optimization of these will have some beneficial effect, but the occupants with a south-facing window will have to resort to frequent use of the blinds/shades to moderate the internal luminous environment. If shaded by a nearby building (Fig. 13b), the daylighting for the south-facing offices could actually be improved if the overshadowing was such that occurrence of direct sun was reduced but the diffuse daylight provision was still adequate. On the north-facing side of the building however, overshadowing could, in the main, lead to a reduced daylighting potential. Other features and design opportunities could serve to improve the daylighting of the space by tempering the direct sun while still admitting sufficient diffuse light. These include brise-soleil (Fig. 13c) and atria (Fig. 13d). The greater the potential for richness/variety in the architectural form, the greater the opportunity for producing a "well-tempered" daylit environment. Low-rise buildings such as schools (Fig. 13e) and also residential buildings (Fig. 13f) offer perhaps the greatest opportunity to realize a "well-tempered" daylit environment since the building mass and the apertures can be designed with the greatest flexibility to control the admittance of direct sun while providing adequate diffuse illumination.

At present, the realization of a "well-tempered" daylit building is more of an art than a science, relying more on the intuition and insight of the experienced daylight designer than what standard evaluative measures such as the daylight factor (see below) can inform.

Guidelines, Measures, and the Evaluation of Daylighting

Illuminance for Task

The absolute levels of illuminance that are needed for any particular task depend on the visual acuity required for the task and, to a lesser degree, the nature of the

environment in which the task is to be carried out. Most developed countries have produced design guides which give recommended illuminance levels depending on task and/or setting. The following is a selection of recommendations produced by the Chartered Institution of Building Services Engineers (CIBSE) [4]:

- 100 lux for interiors used rarely, with visual tasks confined to movement and casual seeing without perception of detail, e.g., corridors, changing rooms, bulk stores, auditoria
- 200 lux for interiors where the visual tasks do not require perception of detail, e.g., foyers and entrances
- 300 lux for interiors where visual tasks are moderately easy, e.g., libraries, sports and assembly halls, teaching spaces, and lecture theaters
- 500 lux for interiors where the visual tasks are moderately difficult and also where color judgment may be required, e.g., general offices, kitchens, laboratories and retail shops
- 1,000 lux for interiors where the visual tasks are very difficult, requiring small details to be perceived, e.g., general inspection, electronic assembly, retouching paintwork, cabinet making, and supermarkets

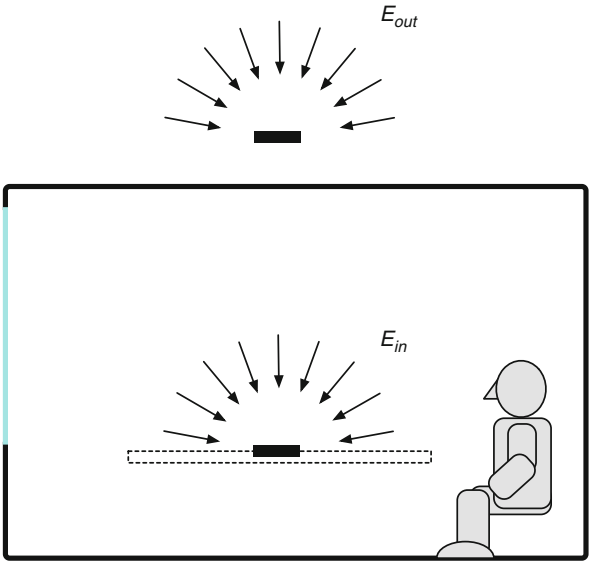
Recommended illumination levels were conceived primarily for the purpose of designing artificial lighting systems and not for the daylighting design of buildings because the variation in the provision of natural daylight is such that it is virtually impossible to deliver specific natural illumination levels without huge fluctuations occurring. For buildings therefore, design guidance was formulated in terms of building properties which are evaluated under a single, static "worst-case" daylight condition: an overcast sky. This is the basis of the daylight factor described in the following section. It is only with recent advances in daylight prediction techniques that absolute levels of daylight illumination under varying sky and sun conditions has become a consideration in the evaluation of the daylighting potential of a building (see section "Climate-Based Daylight Modeling").

The Daylight Factor

Design guidelines worldwide currently recommend daylight provision in terms of the long-established daylight factor (DF) [5]. The daylight factor at a point in an internal space is simply the ratio of internal illuminance E_{in} to unobstructed horizontal illuminance E_{out} under standard CIE overcast sky conditions, Fig. 14. It is usually expressed as a percentage, so there is no consideration of absolute illumination values:

$$DF = \frac{E_{in}}{E_{out}} 100\% \quad (1)$$

The luminance of the CIE standard overcast sky is rotationally symmetrical about the vertical axis, i.e., about the zenith. In other words, the illumination that the standard overcast sky delivers to an internal space will be the same regardless of the compass orientation of the building. And, since the sky is fully overcast, there is no sun. Thus for a given building design, the predicted DF is insensitive to either the building orientation (due to the symmetry of the sky) or the intended locale (since it is simply a ratio). Because the sun is not considered, any design strategies dependant on solar angle, solar intensity, redirection of sunlight, etc., can have no influence on the daylight factor value.



Daylight, Indoor Illumination, and Human Behavior. Figure 14
The daylight factor is the ratio of internal illuminance to unobstructed horizontal illuminance under standard CIE overcast sky conditions

The daylight factor was conceived as a means of rating daylighting performance independently of the actually occurring, instantaneous sky conditions. Hence it was defined as ratio. However, the external conditions still need to be defined since the luminance distribution of the sky will influence the value of the ratio. At the time that the daylight factor was first proposed, it was assumed that heavily overcast skies exhibited only moderate variation in brightness across the sky dome, and so they could be considered to be of constant (i.e., uniform) luminance. Measurements, however, revealed that a densely overcast sky exhibits a relative gradation from darker horizon to brighter zenith; this was recorded in 1901. With an improved, more sensitive measuring apparatus, it was shown that the zenith luminance is often three times greater than the horizon luminance [6]. A revised formulation for the luminance pattern of overcast skies was presented by Moon and Spencer in 1942, and it was adopted as a standard by the CIE in 1955. Normalized to the zenith luminance the luminance distribution of the CIE standard overcast it has the form:

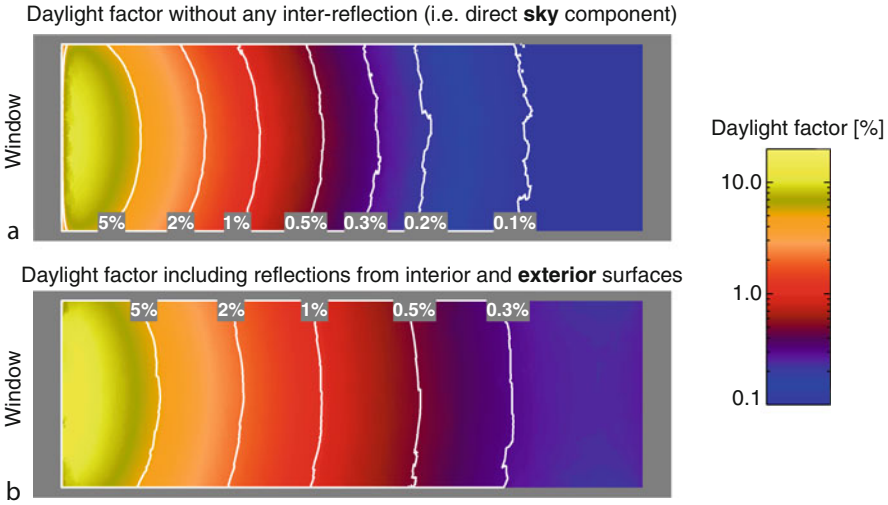
$$L_\zeta = \frac{L_z(1 + 2\cos\zeta)}{3} \qquad (2)$$

where L_ζ is the luminance at an angle ζ from the zenith and L_z is the zenith luminance. Comparisons with measured data have demonstrated the validity of the CIE standard overcast sky model as a representation of dull sky conditions [7]. Thus, since 1955, the daylight factor is strictly the ratio of internal to external illuminance determined under a sky luminance distribution that conforms to the CIE Standard overcast sky pattern (Eq. 2). Note that in papers and reports published prior to 1955, the "daylight factor" is likely to refer to a ratio determined for an actual or assumed uniform sky luminance pattern.

Influence of Building Properties on the Daylight Factor The daylight factor was intended to be a measure of the daylighting potential of a space. The key building properties that determine the magnitude and distribution of the daylight factor in a space are:

- The size, distribution, location, and transmission properties of the windows
- The size and configuration of the space
- The reflective properties of the internal and external surfaces
- The degree to which the external structures obscure the view of the sky

The importance of reflection in a space is illustrated in Fig. 15 which shows the distribution in daylight factor across a plane at desk height (0.8 m) in a rectangular space 3.0 m wide, 9 m deep, and 2.7 m high (i.e., the same space used for the earlier example).



Daylight, Indoor Illumination, and Human Behavior. Figure 15
Example daylight factor

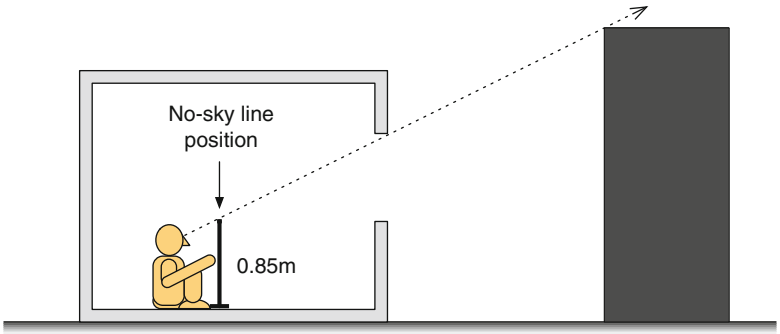
The upper plot shows the distribution in the daylight factor with no inter-reflection of light taking place. This would be the case if all the opaque surfaces had zero reflectance (i.e., perfectly black). This quantity is also known as the direct sky component of the daylight factor since it is entirely due to the sky which is directly "visible" from the point of calculation. The lower plot shows the same distribution but now including the effect of reflection from building surfaces having typical reflectivity values, i.e., 0.7 for the ceiling, 0.5 for the walls, and 0.2 for the floor and external ground. In both cases the daylight factor is greatest closest to the window; however, the proportion of the total that is due to reflected light increases with distance away from the window. At the back of the space, reflected light accounts for more than three quarters of the total daylight factor.

The mean daylight factor across the workplane including inter-reflection is 2.2%. The mean value however is greatly skewed by the high daylight factor values near to the window. The median daylight factor value is only 0.77%, i.e., half the floor area has a daylight factor less than this. One of the measures formulated to determine the evenness of the daylight distribution is called uniformity and it is defined as the minimum daylight factor divided by the mean value for daylight factor. Since the minimum DF value was 0.2%, the uniformity for this space was predicted to be 0.1.

Because the amount of sky visible at the workplane is a governing factor for general illumination, where obstructions are present it is common to estimate the "no-sky line," i.e., that point on the workplane where the sky just ceases to be visible, Fig. 16. If there is no obvious external horizon (i.e., no obstructions outside the window) then, as a rule of thumb, it is common to assume that natural light can penetrate into a space a distance twice that of the floor to ceiling height (for a window that extends from the workplane to the ceiling).

The Contribution of Sun to Daylight The potential for direct sun to enter and illuminate a space is a key consideration in architecture, albeit one that has been traditionally evaluated qualitatively. With scale models this is usually investigated independent of skylight illumination using a simple directional light source called a heliodon to represent the sun, and visually inspecting the resulting shadow patterns across and inside the model. At its most basic, a heliodon could be a small halogen lamp on a movable track. Rather than allowing for a full hemisphere of possible sun positions – and the space it would require – the light source in larger heliodons is usually kept fixed and the scale model is secured to a turntable that is free to rotate in all directions. Thus by positioning the turntable (and attached model) accordingly, it is possible to obtain the correct relative position between the sun (i.e., the light) and the model for a particular time of the day/year [8]. Heliodons have mostly been used by architects and students of architecture.

The shadow pattern approach is essentially qualitative. The brightness of the sun plays no part and light from the sky is not considered. The daylight factor approach (using a sunless sky) and the shadow-pattern approach (using a skyless sun) are essentially incompatible methodologies. Thus it is often difficult to



Daylight, Indoor Illumination, and Human Behavior. Figure 16
Estimating the no-sky line

reconcile the (qualitative) outcome of a shadow-pattern study with a quantitative measure of relative illumination under an overcast sky (i.e., a daylight factor). The inability to meaningfully evaluate daylight in its totality – light from the sun and the sky – in a single, unified schema is believed to be one of the major contributing factors leading to poor design decisions, e.g., overglazing of buildings and the ineffective application of brise-soleil. Computer modeling is used nowadays in preference to scale models to generate shadow patterns, and while this may be more convenient, especially for complicated geometries, the fundamental limitations of the approach are the same.

Determination of the Daylight Factor at the Design Stage

The various methods that can be used to predict daylight factors generally fall into one of the following categories:

- Physical, i.e., measurements in a scale model
- Graphical, tabular, and analytical methods
- Computer simulation, e.g., using a program such as *Radiance* [9]

Physical Modeling Architects have for centuries used physical scale models to study various aspects of building design including natural lighting, and the practice is still commonplace today. Daylight factors were first measured in scale models under actual overcast sky conditions. The measurements of internal and (unobstructed) external illuminance need to be taken simultaneously since the illuminance produced by an actual overcast sky can vary significantly over a period of a minute or even shorter. The daylight factor values obtained under actual conditions are to a fair degree approximations since many seemingly overcast skies have luminance patterns that diverge markedly from the CIE standard description [10]. An artificial sky provides a controlled means of illuminating a scale model for the purpose of taking measurements and also for qualitative appraisal [5]. The most common artificial sky is the "mirror box" design. This has a horizontal sheet of white diffusing material forming the top of the box. The sheet is evenly lit from behind (i.e., from above) by lamps. The four vertical sides of the box are mirrors. These create a sky vault that extends seemingly to infinity on all sides due to multiple reflections between the mirrors. Measurements have shown that the luminance pattern in mirror box skies can approximate that of the CIE overcast sky, and so these can provide a controlled luminous environment for the determination of daylight factors. Many of the larger schools of architecture had artificial skies at one time or another, but they tend to be less used since computer-based methods became more common.

Graphical, Tabular, and Analytical Methods The Waldram diagram is one of several graphical methods that were devised in the early 1900s to predict the direct sky component of illumination under simple sky conditions [5]. The principle of the Waldram diagram is that the half hemisphere of sky visible from a vertical window without obstruction is mapped onto a regular grid such that equal areas of the grid correspond to equal values of direct illumination from the sky. This involves applying a distortion to the representation of actual building features, e.g., the outline of a window, so that they can be shown on the diagram. The Waldram diagram is rarely used nowadays for daylighting evaluation, though some practitioners still resort to it for the arcane practice of determining a "right of light" [11].

Tabular methods such as the BRS tables were introduced in the 1950s and widely used at the time, though now they are only of historical interest. One of the many analytical methods commonly used is the equation to predict the average daylight factor. First proposed by Lynes in 1979 [12], the equation was revised by Crisp and Littlefair in 1984 following validation tests using scale models [13]. The revised equation is:

$$\overline{DF} = \frac{TW\theta M}{A(1 - R^2)} \tag{3}$$

Here T is the effective transmittance of the window(s); W is the net area of side window(s); θ is the angle in degrees subtended in vertical plane by sky visible from the center of a window; M is the maintenance factor; A is the total area of bounding surfaces of an interior, floor + ceiling + walls, including window(s); R is the area-weighted mean reflectance of interior

bounding surfaces. The equation is still used as a rule-of-thumb method at the design stage [14].

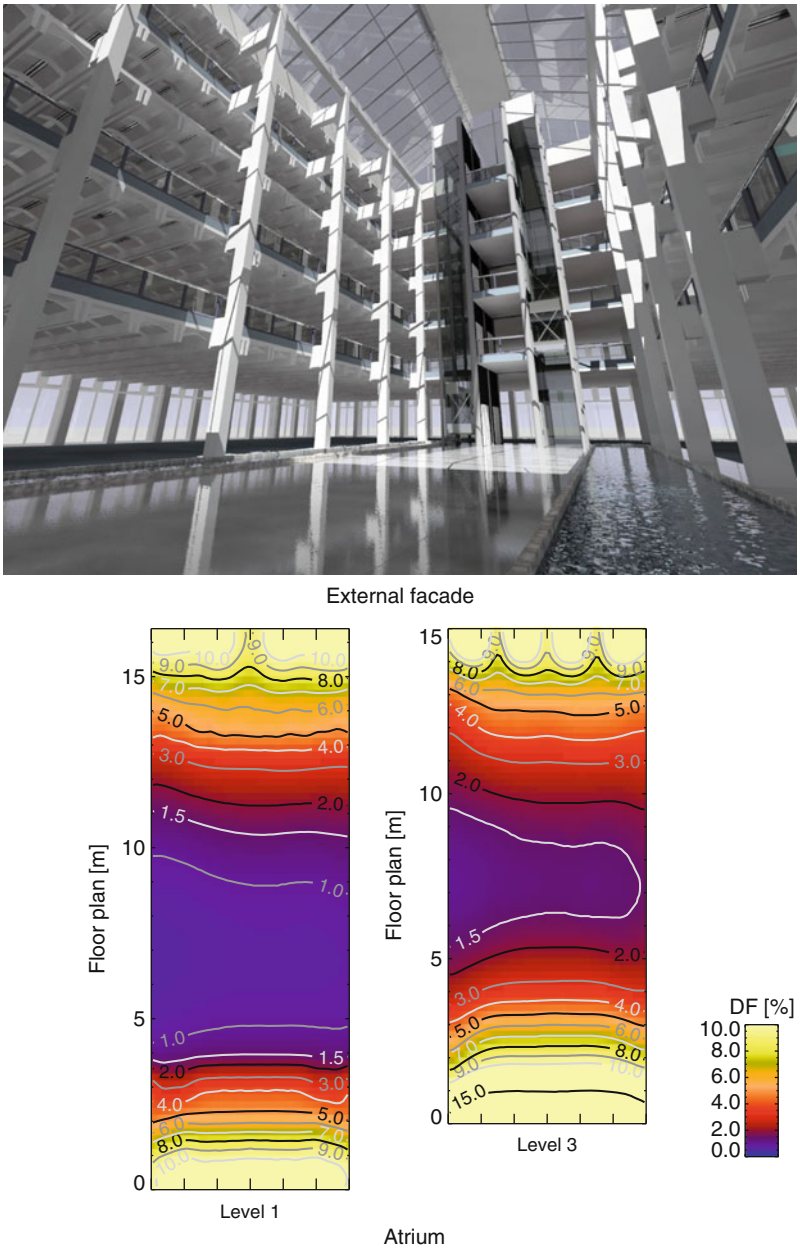
Daylight Prediction by Computer Simulation Computer programs that could predict internal daylight levels in buildings first became widely available in the late 1980s. One of the most commonly used is the freely available *Radiance* lighting simulation system [9]. Lighting simulation programs need to be distinguished from so-called photo-realistic rendering programs such as 3DS Max. The former, which includes *Radiance*, predict the transport of light in a virtual 3D scene using physically based models for the emission, transmission, reflection, and scattering of light. Thus the output can inform on how the actual building might perform, e.g., in terms of visual impression and predicted illuminance levels for a particular sky condition. Photo-realistic rendering in contrast uses various nonphysical means to quickly generate images that give an impression of what the real scene might look like in terms of its underlying 3D geometry, but the shading (i.e., lighting) applied to the surfaces is largely arbitrary. The image in Fig. 17 is a simulation of an atrium space under sunny sky conditions that was computed using *Radiance*. Below the visualization are two plots showing the distribution in predicted daylight factor (i.e., under CIE standard overcast sky) calculated at the workplane height across the floor plans for levels 1 and 3 (Fig. 17).

Accuracy of Daylight Factor Predictions Unlike scale models used to study thermal, acoustic, and structural properties, physical models for daylighting do not require any scaling factors. Thus the daylight conditions in a faithful scale model replica should be the same as for the full-sized building given identical external conditions. This fundamental property of illumination physics proved so compelling to practitioners, and indeed many researchers, that measurements of illuminance in scale models for daylight factor calculation assumed "benchmark" status even though the accuracy had not been rigorously proven.

One of the definitive studies on the accuracy of illumination studies using scale models was that carried out by Cannon-Brookes in 1997 [15]. The study found that the scale model measurements typically exceeded the illuminances in the actual space by 50–100% for overcast conditions and up to 250% for clear skies with sun. These findings, in addition to the inherent problems associated with sky simulator domes [16], have led most practitioners and researchers to choose computer simulation rather than physical models for the bulk of rigorous, quantitative work.

Computer simulation of lighting quantities including daylight has undergone numerous validation tests since the mid-1990s. The degree and exacting nature of these exceed those which physical modeling has been subjected to by a large margin. In particular, the *Radiance* system has undergone more validation studies than any other program, and many of the tests were under daylight conditions in actual spaces. In what is widely considered to be the definitive validation study of a simulation program under real sky daylight conditions, illuminance predictions from *Radiance* were compared with measurements taken in a full-size office space. Measured sky luminance patterns were "mapped" into the simulation program so that the absolute accuracy of the program could be evaluated without the uncertainties that are introduced when the sky luminance pattern has to be estimated using a sky model. The majority of *Radiance* illuminance predictions were shown to be within ±10% of the measured values [17–19]. This and other studies have led to the widespread adoption of *Radiance* by leading consulting engineers. For both researchers and practitioners worldwide, *Radiance* has become the de facto standard for a wide range of applications, not least the evaluation of daylight in buildings [20].

It needs to be borne in mind that validation studies reveal the *potential* accuracy of a daylight simulation program, and that the high accuracy shown in tests can only be approached in "live" projects if strict quality-control procedures are employed. Users of daylight simulation programs such as *Radiance* need to be fairly skilled and, ideally, have a good appreciation of the issues involved, e.g., how to construct a 3D model that is suitable for lighting simulation and how to set to key parameters for the simulation. Studies have revealed that users presented with the same design can produce widely differing predictions for, say, daylight factors, and that suitable training in the use of the simulation tool is vital [21].



Daylight, Indoor Illumination, and Human Behavior. Figure 17
Visualization of atrium under sunny sky conditions and daylight factor plots under the CIE standard overcast sky

Despite the proliferation of computer-based tools for building evaluation and their proven accuracy, scale models still have a great appeal, especially to architects, because they offer an immediate and almost tactile feel for the relation between building geometry, surface properties, and the character of illumination. This is especially apparent when the spatiotemporal dynamic play of light in a model is observed using a heliodon to mimic the progression of sunlight through the day.

The Daylight Factor and Standards

Codes, standards, and guidelines for daylighting are invariably advisory rather than mandatory. As noted

earlier, codes, etc., worldwide are almost all founded on the daylight factor. One of the key standards documents for daylight in buildings is the British Standard 8206–2 *Lighting for buildings – Part 2: Code of practice for daylighting* [22]. That document gives the following recommendation:

- It is considered good practice to ensure that rooms in dwellings and in most other buildings have a predominantly daylit appearance. In order to achieve this, the average daylight factor should be at least 2%. If the average daylight factor in a space is at least 5% then electric lighting is not normally needed during the daytime, provided the uniformity is satisfactory. If the average daylight factor in a space is between 2% and 5%, supplementary electric lighting is usually required.

For uniformity, the same standard advises that

- . . . the minimum illuminance on a particular task area should not fall below 0.7 times the average illuminance on that task area.

This indicates that, for the 2% average criterion, the minimum daylight factor should be no less than 1.4%. It is instructive to compare this advice with the daylight factor distribution for a deep-plan office space shown in Fig. 15b. Although the mean daylight factor meets the requirement of 2%, the uniformity value for the space was only 0.1, much lower than the recommended minimum of 0.7. The median value of just below 0.8% better indicates just how much of the space falls below the minimum value of 1.4%. This simple example shows that there are quite fundamental limitations to the depth of daylight (strictly, *daylight factor*) penetration that can be achieved using only vertical glazing from one direction. Strategies to extend the penetration of daylight from perimeter windows include having large floor to ceiling heights with tall or clerestory windows (see section on "Daylighting Strategies"). If the window receives significant sun, then a light-shelf (Fig. 6) might serve the dual purpose of shading near to the window and redirecting sunlight deeper into the space. Note, however, that the effectiveness of a light-shelf design cannot be directly assessed using the daylight factor since neither sunlight nor sunny sky conditions figure in the evaluation.

Daylight, Indoor Illumination, and Human Behavior.

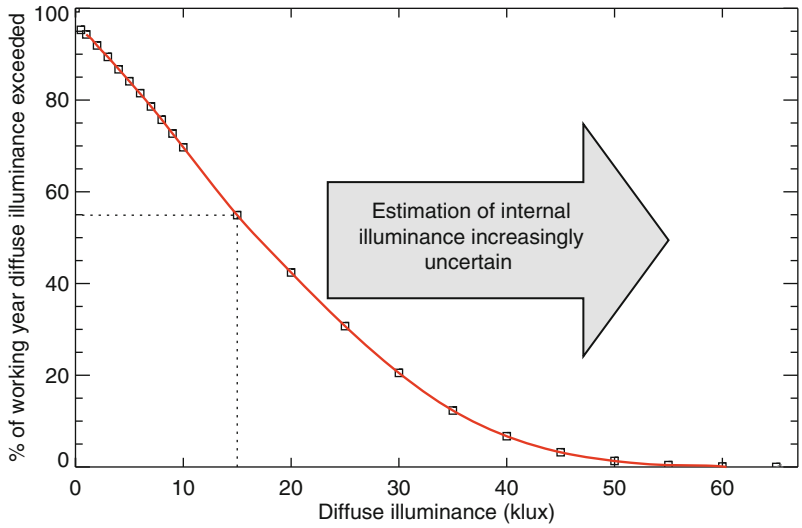
Table 1 Typical daylight factor values recommended in guidelines during the latter half of the twentieth century

Nondomestic buildings	
Church	1% minimum
Factory	5% minimum
Office	2% minimum
Classroom	2% minimum
Hospital ward	1% minimum
Dwellings	
Bedroom	0.5% at 3/4 of room depth
Kitchen	2% at half of room depth
Living room	1% at half of room depth

Recommended Daylight Factor Values A set of daylight factor values as might be typically recommended in various guideline documents are presented in Table 1. Historically, the recommended values for factories have been high since these were often large-area, single-story buildings which could be very effectively daylit by rooflights (invariably configured to avoid the ingress of direct sun).

Deriving Absolute Values from Daylight Factors It is possible to convert a daylight factor value (i.e., a relative measure of illumination under a static sky condition) into an estimation of the annual provision of daylight using cumulative diffuse illuminance availability curves. An example curve based on measurements taken at Kew (near London, UK) is shown in Fig. 18. This curve gives the percentage of the year for which a diffuse horizontal illuminance is achieved during, say, normal working hours – in this case 09:00 to 17:30. Applying a simple technique, cumulative internal illuminance availability can be estimated from daylight factor values and the curves (or similar charts) of cumulative diffuse sky illuminance [23]. This gives a first-order approximation to annual daylighting provision from which supplementary lighting requirements can be estimated.

For example, suppose that the minimum required internal illuminance at a point in an office is 500 lux, and that a daylight factor evaluation using the CIE



Daylight, Indoor Illumination, and Human Behavior. Figure 18
Cumulative diffuse illuminance availability, Kew, UK

standard overcast sky (equation, scale-model or simulation) predicts a daylight factor value of 3.3%. The minimum diffuse sky illuminance which provides an average internal illuminance of 500 lux is therefore:

$$\frac{500 \times 100}{3.3} \approx 15,000 \qquad (4)$$

It can be determined from Fig. 18 that a diffuse sky illuminance of 15,000 lux is exceeded for about 55% of the normal working time. The CIE standard overcast sky is likely to be a reasonable approximation to some of the duller skies in the cumulative distribution. However, only about 40% of the skies in the Kew climate file for the UK can be classed as heavily overcast [18]. For locales sunnier than the UK, the percentage of overcast skies throughout the year will of course be lower. The fundamental weakness of this approach is that it tries to extend the daylight factor modeling paradigm (i.e., relative illumination under standard overcast sky conditions) to somehow account for the illumination effect of non-overcast skies. Since real non-overcast skies diverge enormously from overcast sky luminance patterns, the estimations of illuminance from the non-overcast skies in the distribution will be greatly in error. Additionally, the method cannot of course account for the illuminance contribution of the sun – either directly or from reflection. Thus the divergence between estimation and reality increases as higher illuminances in the cumulative distribution are considered (Fig. 18). Needless to say, this approach is highly inappropriate for building designs where the redistribution of direct beam radiation to provide diffuse illuminance is a significant feature of the daylighting system, as is the case with designs that make use of deep window reveals, light shelves, light wells, etc. Indeed, redirection of direct beam illumination will occur in all buildings to a greater or lesser degree even when it is not an explicit design feature. Despite the limitations, the cumulative illumination approach must be considered to be an advance on the daylight factor alone because the results, however flawed, do at least depend on absolute measures of the climate for the locale. In other words, the results for St. Petersburg will be different to those for Cairo. The effect of building orientation on internal illuminance, i.e., the difference between north- and south-facing glazing, can be roughly estimated by applying so-called orientation factors [24]. However it is believed that this is hardly ever used by practitioners.

Daylight Guidelines Post 2000 Since the year 2000, the role that daylight evaluation plays in the design process has acquired a new impetus as the need to demonstrate compliance with various "performance indicators" becomes ever more pressing. Two of the most used rating systems are BREEAM (The BRE

Environmental Assessment Method) and LEED (Leadership in Energy and Environmental Design) which originated in the UK and USA, respectively, though they are both used worldwide [25, 26]. Both the LEED and the BREEAM websites chart the growth in the building projects that have been certified using the respective schemes. These and similar rating systems are actively promoted by government departments and lobby groups. As a consequence, building designers are resorting more and more to prediction methods (invariably simulation) as a means of demonstrating compliance with the various schemes [20]. This, one might reasonably hope, would lead to noticeable improvements in the practice of design evaluation, which in turn should improve the likelihood of realizing a well-daylit building. However, basing design evaluation on the daylight factor has not been proven to result in better daylit buildings.

The Daylight Factor and Actual Daylighting Conditions

The daylight factor was formulated long before the computation of actual illumination levels became a practical possibility. Thus the simplifications inherent in the formulation were, back then, a necessary expediency. As noted, a major issue with the daylight factor is that actual daylight illumination conditions deviate markedly from that described by the overcast sky paradigm. This is so even for Northern Europe where there is a commonly held belief that skies are "mostly" overcast and so use of the daylight factor as a basis for evaluation is justified. A paper by Littlefair in 1998 gives annual cumulative internal illuminance measurements for a point in similar rooms with north- and south-facing glazing [27]. The rooms were unshaded and unoccupied. An illuminance of 200 lux was achieved for approximately 58% and 68% of the year for the north- and south-facing spaces respectively. However, an illuminance of 400 lux was achieved for only 12% of the year for the north-facing space with more than four times that occurrence (51%) for the south-facing space. Of course, for sunnier climates the effect of orientation on daylight illumination will be greater still.

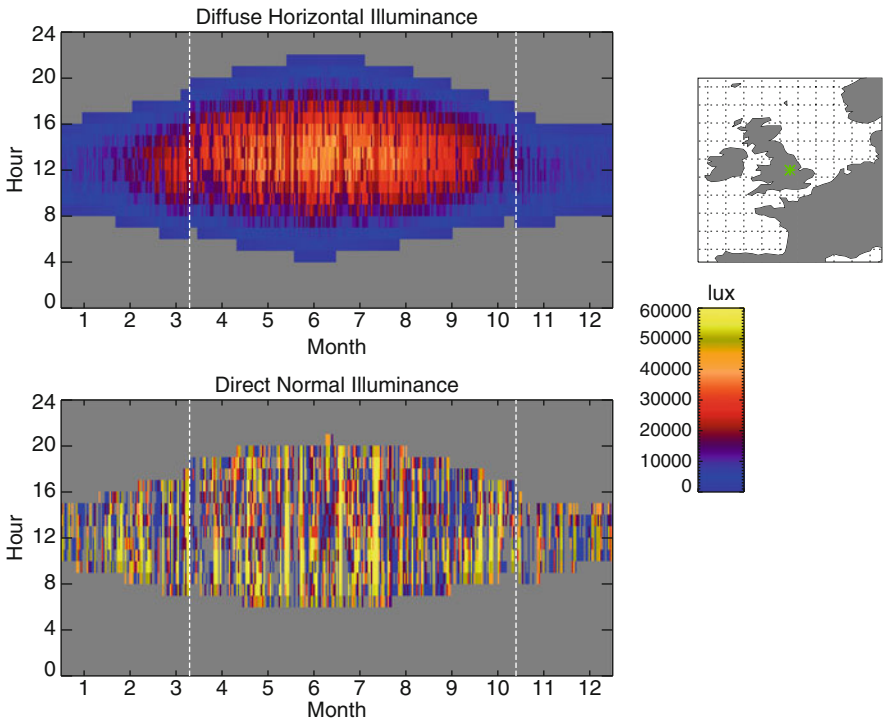
An unfortunate consequence of the long-standing and often uncritical use of the daylight factor is that the terms "daylight" (as defined in the Introduction) and "skylight" are often used interchangeably. This leads to confusion where precise definitions are required. Some of this muddle has resulted from the conflation of "daylight" per se with what is predicted by the daylight factor. For example, expressions such as "the daylight factor was used to evaluate daylight levels" are common in both research and practice literature. The daylight factor is precisely what it was defined to be: a ratio of illuminances under a specific sky condition. The daylight factor therefore is, in reality, a proxy for actual daylight illumination. Thus, what the daylight factor communicates is in fact very different from the actual illumination levels that result from the full range of naturally occurring sun and sky conditions.

Extending the basis of the daylight factor approach by incremental means has proved problematic. It is a straightforward matter to use in a daylight simulation non-overcast sky conditions, e.g., the CIE clear sky luminance pattern with sun (see Fig. 17). To be useful for evaluation purposes however, the luminous output of the sun and sky must be known since absolute values and not ratios must be considered. Extending the daylight factor notion of ratios to non-overcast skies with sun results in essentially meaningless values and should be avoided. When absolute values for luminous quantities are predicted (e.g., lux at the workplane) then the values used to normalize the output from the sun and sky must be justified, e.g., diffuse horizontal illuminance for the sky and direct normal illuminance for the sun. Ideally, these should be based on typical values for, say, a clear, sunny day in summer. The predicted quantities however will be of very limited value for any estimation of prevailing daylight levels in the building since they are indicative only of conditions for particular sun and sky conditions occurring at a particular time of the day/year. In other words, such an evaluation would offer merely a single "snapshot" of the multitude of naturally occurring daylight conditions due to all the possible combinations of sun and sky conditions occurring at various times throughout the year. Estimating overall daylighting performance from snapshot evaluations could be highly misleading. The parameters governing the availability of daylight do not lend themselves to any form of averaging. While it can be informative to determine, say, a monthly average for a scalar quantity such as temperature, illumination is

strongly dependent on the directional character of the incident light. Associated with every non-overcast sky and sun condition are the solar altitude and azimuth which, of course, vary continuously throughout the day. In terms of providing a basis for predicting measures of illumination, the notion of "average" days is less than useful because an "average" sun position would give entirely misleading patterns of illumination. Some guidelines (e.g., LEED and ASHRAE) have allowed so-called clear sky options where a specified daylight illuminance must be achieved (usually at the workplane) for clear sky conditions at a particular time, e.g., at noon on the equinox. These approaches are problematic for a number of reasons. In particular, some of them provide no guidance on normalizing the sky output, a significant omission since absolute values are now the target. Furthermore, the method seems to recommend evaluation under clear sky conditions *without* a sun – a physically impossible illumination condition in nature.

The true nature of illumination from the sun and sky for any particular locale can only be appreciated by examining the luminous output from both the sun and the sky over a period of a full year. The principal sources of annual climate data are the standard weather files which were originally created for use by dynamic thermal modeling programs [28]. These datasets contain hourly averaged values for a full year, i.e., 8,760 values for each parameter. The key daylight parameters stored in the weather files are the diffuse horizontal illuminance and the direct normal illuminance. The diffuse horizontal illuminance is the visible light energy from the unobstructed sky that is incident on a horizontal surface. The direct normal illuminance is the visible light energy from the sun that is incident on a surface which is kept normal to the beam of radiation, i.e., the photocell always "points" directly to the sun. If the climate file contains only irradiance (i.e., total energy values), then it is converted to illuminance using a luminous efficacy model [29].

A visualization of the illuminance data from a standard weather file is given in Fig. 19. The time-series data of 8,760 values has been rearranged into an array of 365 days (x-axis) by 24 h (y-axis). The shading at each hour



Daylight, Indoor Illumination, and Human Behavior. Figure 19
Illuminance data from the CIBSE standard climate file for Nottingham

indicates the magnitude of the illuminance – see legend – with zero values shaded gray. Presented in this way it is easy to appreciate both the prevailing patterns in either quantity and their short-term variability. Most obvious is the daily/seasonal pattern for both illuminances: short periods of daylight in the winter months, longer in summer. The hour-by-hour variation in the direct normal illuminance is clearly visible, though it is also present to a lesser degree in the diffuse horizontal illuminance (i.e., light from the sky). The data is for Nottingham, UK, the location of which is identified in the map above the legend. Local time is shown, i.e., summertime is local time plus one hour. The start and end period of summertime are indicated by vertical dashed lines in each of the figures. Recall the patterns of indoor illumination for the four components of daylight given in Fig. 3. Each of the 4,380 (i.e., the daylight hours) unique combinations of sky and sun conditions in the weather file (Fig. 19) will result in a unique pattern of internal daylight illumination.

Both diffuse and direct illuminances will, in reality, vary over periods much shorter than an hour. Interpolation of the dataset to a time-step shorter than one hour will provide a smoother traversal of the sun, which may be necessary when using the data for simulation of daylight. Interpolation alone, however, will not introduce short-term variability into the values for diffuse horizontal and direct normal illuminance. If required, this variability would have to be synthesized using stochastic models [30].

The illuminance data in the standardized weather files reveal the true nature of the patterns in daylight illumination from the sun and the sky. It is also evident from the visualization of the data (Fig. 19) that any "snapshot" evaluation using just part of the data would not be representative and could lead to highly flawed conclusions regarding the daylighting performance of the building. How these data might be used in their entirety to better predict actual building performance is described in the section on "Climate-Based Daylight Modeling."

Human Factors and Daylight

Daylight and Visual Comfort

A good provision of daylight is now considered to be highly desirable in terms of improving occupants' well-being and productivity [31, 32]. Daylight, however, can cause visual discomfort by inducing glare and veiling reflections. Efforts to control glare often result in the loss of predicted daylight benefit as occupants deploy blinds, etc., which may remain closed long after the glare condition has diminished.

In the CIBSE Lighting Guide LG7, glare is defined as a "Condition of vision in which there is discomfort or a reduction in the ability to see details or objects, caused by an unsuitable distribution or range of luminance, or to extreme contrasts" [33]. There are two types of glare: disability glare, where stray light reaching the eye results in a reduction of visibility and visual performance, and discomfort glare, which leads to users' discomfort, often with less immediately noticeable effects such as headaches or posture-related aches after work. Glare can be caused by direct sunlight through a window or by the luminance differences between bright areas such as windows with bright sky views and the darker task area. Furthermore, veiling reflections on reflective surfaces such as computer screens can affect visual comfort at workstations facing away from the window.

The majority of office workers now use (vertical) computer screens during all or part of their workday [34]. They are far more likely to be affected by glare from windows than their counterparts of 10 or 20 years ago who were largely occupied with paper-based tasks on horizontal surfaces [35]. It is considered an imperative, therefore, to attempt to minimize glare when designing for daylight provision in offices. Current international and European standards regarding visual comfort in the office environment aim to address this issue (EN ISO 9241–7:1998, EN 12464–1:2002, EN 29241–3:1993). A number of guidelines have been published which aid designers and architects in their efforts to reduce glare in order to achieve comfortable visual environments, such as the CIBSE Lighting Guides LG3 and LG7, the Code for Lighting [36], and CIBSE Guide A [4].

While there are accepted, albeit imperfect, models for the potential glare effect of (fixed output) luminaires, it is recognized that glare from daylight sources is poorly understood [37]. The first daylight glare formulations were extrapolations from studies of discomfort glare due to artificial lighting [38]. The light sources used in those studies subtended relatively

small solid angles from the viewpoint of the subject, and the luminance conditions (source and environment) were very different from typical daylit offices. In short, those extrapolations proved to be inadequate for the purpose of determining discomfort glare from daylight. There have been a number of attempts to improve glare formulations for use in daylight situations, such as Daylight Glare Index (DGI), Unified Glare Rating (UGR), and the New Daylight Glare Index (DGIN), but the problem persists. A review in 2005 by the chair of the International Commission on Illumination (CIE) Technical Committee on glare concluded that the "available assessment and prediction methods are of limited practical use in daylit situations" [37].

Numerous field studies have been carried out in order to investigate glare from daylight through windows. The earliest discomfort glare studies used large-area artificial light sources to mimic the effect of daylight through windows [5]. However, the sensation of discomfort that may arise from staring at a featureless (artificial) light source seems to be very different from that produced by viewing a comparably high-luminance natural scene [39, 40]. Moreover, as outlined in a recently published review of research regarding occupant satisfaction with the luminous environment, there are multiple additional factors that affect occupants' visual comfort, such as preferred light levels and access to lighting and shading controls [32]. Studies which have included occupant surveys in combination with controlled measurements of the luminance were often set in laboratory environments quite different from the usual workplace.

Considering that the natural variability in daylight contains seasonal, daily and short-term components and that discomfort glare is known to depend strongly on directional factors as well as the scalar magnitude and distribution of the luminous field, findings from existing studies have limited use when describing glare conditions experienced in real office spaces. Value judgments appear to be the only ground on which discomfort from glare may be assessed, but it is nevertheless essential that these subjective assessments are linked to objective and quantifiable data or phenomena [37]. New approaches to assessing glare that may overcome the current limitations in understanding are described in the section "New Approaches to Measuring the Daylit Environment."

Daylight and Health/Productivity

The primary concern in the daylighting of buildings has generally been to provide illumination for task, e.g., 500 lux on the horizontal workplane. However, in the last few decades there has been a gradual increase in awareness of the nonvisual effects of daylight/light received by the eye [41]. (We exclude from consideration here skin exposure effects such as tanning and the production of vitamin D.) It is well known that building occupants almost without exception will prefer a workstation with a view of the outdoor environment to a windowless office [42]. A view to the outside indicates of course the presence of daylight, although the relation between view and daylight provision is not straightforward, being dependent on many factors. Might there be productivity and well-being benefits in providing building occupants with well-daylit spaces? In addition to subjective preferences for daylit spaces, it is now firmly established that the light has measurable biochemical effects on the human body, in particular with respect to maintaining a healthy sleep-wake cycle. Could the quality and nature of the internal daylit environment have a significant effect on the health of the human body which can be proven through the measurement of, say, hormone levels? Evidence is suggestive of links between daylight exposure and both health and productivity; however, there is insufficient knowledge at present to conflate these two effects and so they are discussed separately in the following sections, beginning with health.

Health The daily cycle of day and night plays a major role in regulating and maintaining biochemical, physiological, and behavioral processes in human beings. This cycle is known as the circadian rhythm – the term "circadian" comes from the Latin circa, "around," and diem or dies, "day," meaning literally "approximately one day." Circadian rhythms occur in almost all organisms from bacteria to mammals. The circadian rhythm is endogenous meaning that it is produced from within the organism, i.e., what is commonly referred to as the "body clock." However, for many organisms the cycle

needs to be adjusted or entrained to the environment by external cues, the primary one of which is daylight.

The primary circadian "clock" in mammals is located in the suprachiasmatic nucleus (or nuclei) (SCN), a pair of distinct groups of cells located in the hypothalamus. The SCN receives information about illumination through the eyes. The retina of the eye contains not only the well-known photoreceptors which are used for vision (i.e., rods and cones) but ganglion cells which respond to light and are called photosensitive ganglion cells. The SCN in turn coveys signals to the pineal gland, which, in response, controls the secretion of the hormone melatonin. Secretion of melatonin peaks at night and ebbs during the day, i.e., its presence modulates the wake/sleep patterns.

The failure to maintain a circadian rhythm that is firmly entrained to the natural 24-h cycle of daylight results in many negative health outcomes for humans, though not all are fully understood. The degree and severity of the outcomes usually depends on the period over which the cycle is disturbed. A transitory disturbance to the circadian cycle familiar to many who have experienced a long-haul flight is jet lag. When traveling across a number of time zones, the body clock will be out of synchronization with the destination time, as it experiences daylight and darkness contrary to the rhythms to which it has grown accustomed. Depending on the individual, it can take a few days to reset the body clock to the local day-night cycle.

Less immediately obvious in its effects than jet lag is the chronic persistence of a poorly entrained circadian rhythm. This was first noticed in shift-workers; however, it is believed to be one of the factors in the increasing occurrence of sleep-disturbance and related conditions in the wider population of the developed world [43]. While the symptoms of sleep-disturbance can be at first mild, e.g., sleepiness, fatigue, and decreased mental acuity, the long-term persistence of the condition may result in significant impacts on both health and worker productivity.

The duration, intensity, and spectrum of the light received at the eye are the principal factors determining the degree of nonvisual effect, and thus a key factor in the entrainment of the circadian cycle. Another important factor is the time of day when the light is applied. Compared to the luminous efficiency function of the eye which has a peak value at 555 nm, the action spectrum for the suppression of melatonin is known to be shifted to the blue end of the spectrum and has a peak around 450–480 nm [44]. The relative suppression of melatonin as a function of light intensity and color temperature for artificial lighting was determined by Mclntyre et al. in 1989 [45]. The illuminances at the eye required for the effective suppression of melatonin are of the order of 1,000 lux depending on the spectrum. Note that the vertical illuminance at the eye is typically one fifth that delivered to the horizontal workplane by (overhead) artificial lights. Thus, a workplane illuminance of around 5,000 lux (much higher than the typical design levels of 300–500 lux) would be needed to provide 1,000 lux at the eye. Thus it is argued that interior lighting levels, as currently recommended and practiced, may be insufficient for circadian regulation [46]. Daylight often provides illuminances significantly higher than the design level, though this is only in close proximity to windows and perhaps also highly daylit spaces such as atria. If the typical illuminances in these zones are high – but not so great that blinds are needed – then those building users that regularly occupy the well-daylit spaces may perhaps experience stronger and more regular circadian entrainment stimuli than those users away from windows who are habitually exposed to lower illuminance at the eye level. These considerations have resulted in the notion that a building through its daylighting may possess a *circadian efficiency* [47]. However, such assertions are presently highly speculative and should be treated as such until the evidence becomes more compelling [48].

There is some evidence that daylight exposure can affect postoperative outcomes in patients and, consequently, that daylight should be a consideration in hospital design. One of the key studies in the literature is that by Walch et al. on patients recovering from elective cervical and lumbar spinal surgery. The patients were housed on either the "bright" or "dim" side of the same hospital unit. The study determined that sunlight exposure was associated with both improved subjective assessment of the patients and also reduced levels of analgesic medication routinely administered to control postoperative pain [49].

Productivity and Performance Bright lighting is generally believed to make people more alert, and

well-daylit spaces are generally perceived by occupants to be "better" than dim gloomy ones. The link between objective measures of productivity and the quality of the daylit environment is however elusive because worker productivity is influenced by numerous factors, and it is proving challenging to isolate the effect of just daylight. Furthermore, it is generally not a practical option to monitor the long-term daylight exposure for hundreds of subjects using light meters. Thus the daylighting potential is often estimated from space parameters that can be easily assessed by a site visit. This adds another measure of uncertainty to these studies since daylight exposure itself is a highly variable quantity. Some reports are mostly anecdotal, e.g., good daylighting has been identified as a factor in staff retention [50].

Perhaps the best known studies that have attempted to link productivity to daylight are those carried out by the Heschong-Mahone Group (HMG) in California, USA. The HMG schools' study claimed that

- ... students with the most daylighting in their classrooms progressed 20% faster on math tests and 26% on reading tests in one year than those with the least daylight. Similarly, students in classrooms with the largest window areas were found to progress 15% faster in math and 23% faster in reading than those with the least window areas. Students that had a well-designed skylight in their room, one that diffused the daylight throughout the room and allowed teachers to control the amount of daylight entering the room, also improved 19% to 20% faster than those students without a skylight [51].

Other studies by HMG have made similarly compelling claims in improved productivity for office workers and larger retail sales in well-daylit compared to poorly daylit spaces [52, 53]. However, until independent studies corroborate these reports, the findings must be considered suggestive of an effect rather than conclusive proof of a strong relationship between daylight and productivity/performance [54].

It should be noted that improved productivity has also been claimed for exposure to higher than usual levels of artificial lighting. In many regards, it is easier to isolate the effects of short periods of enhanced artificial lighting than it is for long periods of (highly variable) daylight. A Japanese study found that bright lighting in the office (2,500 lux compared to 750 lux, provided for 2 h in the morning and 1 h after lunch for several weeks) boosted alertness and mood, especially in the afternoon. It also seemed to promote melatonin secretion and fall in body temperature at night, changes that should improve the quality of sleep. Although this work was based on a small number of people and further work is needed, it shows promise for alterations in office lighting in terms of productivity and health of the workers [55]. These findings have led to developments in the use of artificial lighting where controlled exposure across the entire floor-plate of a building is a straightforward matter [56].

Daylight and the User Operation of Lights and Shades

Almost all the commonly used spaces in side-lit buildings will have some form of shading device, e.g., venetian blinds, that is either user-deployed or automatically controlled. Even if the fixed architectural form offers effective shading from direct sun much of the time, there are, nevertheless, likely to be occasions when blinds will need to be used to block either direct sun and/or the view of bright patches of sky. Blinds are usually formed either from individual sections called slats or are continuous. Slatted blinds can have either horizontal or vertical slats.

Users entering a space where there is little daylight will of course switch on the electric lights. The probability that users switch on electric lights was found to be correlated with the minimum daylight illuminance on the working plane [57]. The correlation presented by Hunt in 1980 was based on just a handful of samples and there was considerable scatter in the switch-on probability when the daylight illuminance was in the range 50–500 lux, which is typical of the range experienced in many buildings. A later study provided support for the Hunt model, but as with the original study there was large scatter in the measured daylight illuminances that triggered the switching on of lights [58]. In addition to the switch-on probability, there will also be switch-off probabilities. Relatively little field study data has been published regarding switch-off behavior, and determining a correlation with daylight is more confounding than for switch-on since other factors come into play. For example, switch-off probabilities

could be significantly determined by the overall appearance of the space and the particular design of the lights, since it is sometimes not obvious to the occupant that lights have been left on when daylight provision is high.

In an effort to account for the variability in occupants' light switching behavior, stochastic models were introduced in 1995 by Newsahm et al. [59]. This means that whenever a (simulated) user is confronted with a control decision, i.e., to switch on the lighting or not, a probabilistic (i.e., stochastic) process is initiated that determines the outcome of the decision. This approach was further refined by Reinhart in a model designated "Lightswitch-2002," which provides a unified modeling framework for both the switching on and off of electric lights and the operation of blinds/shades to control direct sun [60]. A key input parameter for the Lightswitch-2002 is the user type, of which there are four in the model. The type determines the degree to which the simulated user attempts to make the most of the daylight contribution by both optimizing the blinds and switching off unnecessary artificial lighting, and the results for, say, predicted energy use for lighting are highly sensitive to the assumed type.

Recent Advances in Daylighting

The consideration of daylight in buildings has recently undergone a radical reevaluation. For much of the latter half of the twentieth century, the "objective" evaluation of a building design relied almost exclusively on the daylight factor (section "The Daylight Factor"). Of arguably greater value than a daylight factor analysis was the advice of an experienced daylight designer. Although they commonly make use of the daylight factor, the real value of the designer's expertise however is in envisioning those many aspects of daylight provision that are *not* accounted for by the daylight factor. These aspects are many and varied. Key amongst them, however, are the contribution of the sun to the overall illumination of the building and the potential for glare resulting from direct sun and/or skylight. The first of these – the illumination contribution of the sun – can only be very approximately estimated. In truth, it is a qualitative judgment founded on experience and intuition rather than numerous computations of light transfer. The second depends in part on a consideration of geometrical relations between the progression of the sun and the configuration of the building, i.e., the windows of the building, their orientation, and any nearby obstructions. This involves envisioning the progression of the sun-illuminated surfaces inside the building, and estimating the potential for views of bright sky that might be a cause for glare. In other words, for either case there is an envisioning of sorts by the daylighting expert of the spatio-temporal dynamics of daylight illumination. These evaluations can be informed to a limited degree by shadow pattern studies of solar penetration. In addition, of course, an experienced designer will offer advice on a great many other, secondary aspects of daylighting design for the building. However valuable the advice offered by the daylight designer, it is not something that could be distilled into a codified scheme and, ultimately, some numerical measure of predicted performance. If daylighting experts were not part of the design team, then a routine daylight factor evaluation accompanied perhaps by a shadow-pattern study would be the sum total of the "daylighting evaluation." Thus a completed building with good daylighting was more likely to be a product of chance and good fortune than design. Inevitably, continued reliance on the half-century-old daylight factor led also to a sense of stagnation in sectors of the daylighting research community.

Two seemingly concurrent, but out-of-step and totally independent, developments have changed both the perceived importance and the nature of daylight evaluations. The first is the increasing demand to demonstrate compliance at the design stage with recommended measures of building performance, e.g., the LEED rating system. The need for this appears to be widely accepted throughout the developed world, and the rate of uptake by practitioners is ever increasing in response to pressure and encouragement from governments, regulatory bodies, etc. For those striving to effect good daylighting design, however, the race for compliance is by no means entirely good news because the recommendations are founded on schema such as the daylight factor that ignore fundamental parameters such as building orientation and prevailing climate. The second development is a major advancement in the way that daylight evaluations are carried out. This advancement, called climate-based daylight modeling, can address the very real concerns that practitioners and researchers are now voicing regarding the high

potential for "compliance chasing" resulting in poor design choices for buildings [61]. An important addition to these developments is the emergence of an array of new facade and glazing technologies for improving the daylighting of buildings, both new and existing. And in the last five years, camera-based measurement techniques that can characterize the luminous environment with an unprecedented level of coverage have become available to everyday practitioners. These developments, recently established and emerging, are described in the following sections.

Climate-Based Daylight Modeling

Climate-based daylight modeling is the prediction of various radiant or luminous quantities (e.g., irradiance, illuminance, radiance, and luminance) using sun and sky conditions that are derived from standardized annual meteorological datasets. Climate-based modeling delivers predictions of absolute quantities (e.g., illuminance) that are dependent both on the locale (i.e., geographically specific climate data is used) and the fenestration orientation (i.e., accounting for solar position and nonuniform sky conditions), in addition to the space's geometry and material properties. The operation of the space can also be modeled to varying degrees of precision depending on the type of device (e.g., luminaire and venetian blinds) and its assumed control strategy (e.g., automatic, by occupant, or some combination). The computational overhead and complexities introduced when attempting to model the operation of the space are discussed later.

The term "climate-based daylight modeling" does not yet have a formally accepted definition – it was first coined by Mardaljevic in the title of a paper given at the 2006 CIBSE National Conference [62]. However, it is generally taken to mean any evaluation that is founded on the totality (i.e., sun and sky components) of time-series daylight data appropriate to the locale over the course of a year. In practice, this means sun and sky parameters found in, or derived from, the standard meteorological data files which contain 8,760 hourly values for a full year. Given the self-evident nature of the seasonal pattern in sunlight availability, a function of both the sun position and the seasonal patterns of cloudiness, an evaluation period of 12 months is needed to capture all of the naturally occurring variation in conditions that is represented in the climate dataset. It is also possible to use real-time monitored weather for a given time period, if calibration to actual monitored conditions within a space is desired. Standardized climate datasets are derived from the prevailing conditions measured at the site over a period of years, and they are structured to represent both the averages and the range in variation that typically occurs. Standard climate data for a large number of locales across the world are freely available for download from several websites. One of the most comprehensive repositories is that compiled for use with the EnergyPlus thermal simulation program [63]. This contains freely available climate data for over 1,200 locations worldwide.

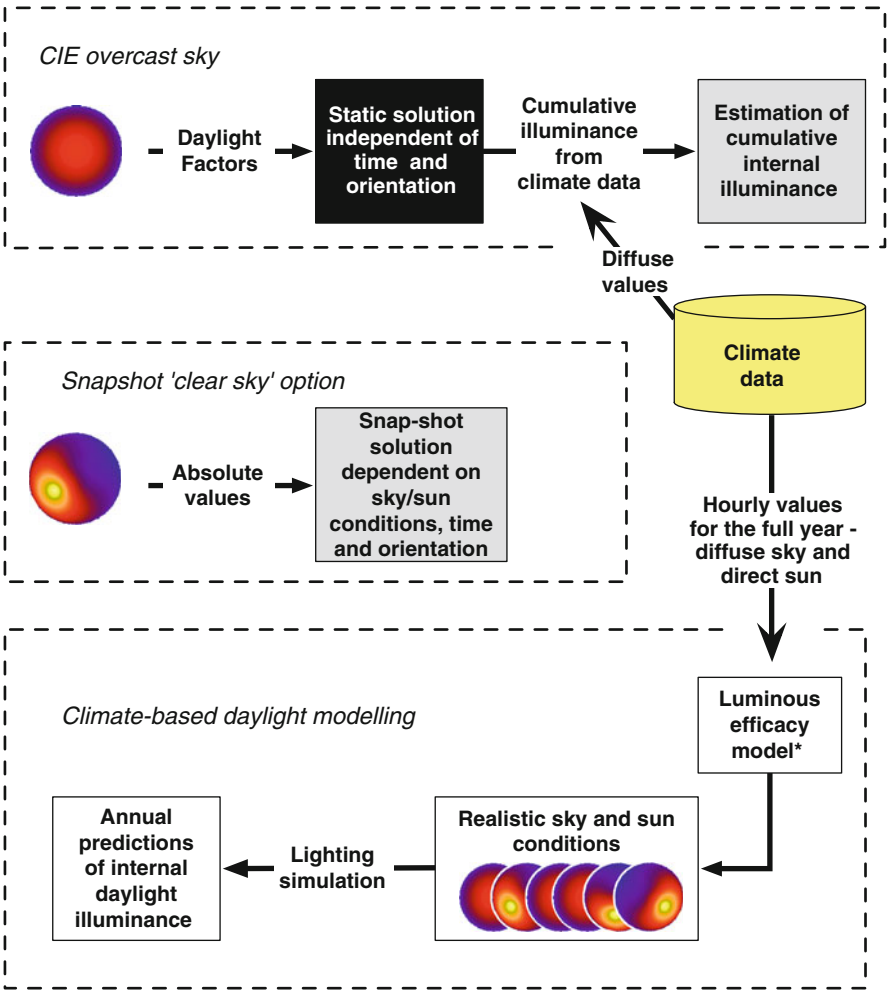
There are a number of possible ways to use climate-based daylight modeling [64–68]. The two principal analysis methods are cumulative and time-series. A cumulative analysis is the prediction of some aggregate measure of daylight (e.g., total annual illuminance) founded on the cumulative luminance effect of (hourly) sky and the sun conditions derived from the climate dataset. It is usually determined over a period of a full year, or on a seasonal or monthly basis, i.e., predicting a cumulative measure for each season or month in turn. Evaluating cumulative measures for periods shorter than one month is not recommended since the output will tend to be more revealing of the unique pattern in the climate dataset than of "typical" conditions for that period. The cumulative method can be used for predicting the microclimate and solar access in urban environments, the long-term exposure of art works to daylight, and quick assessments of seasonal daylight availability and/or the requirement for solar shading at the early design stage. Time-series analysis involves predicting instantaneous measures (e.g., illuminance) based on each of the hourly (or sub-hourly) values in the annual climate dataset. These predictions are used to evaluate, e.g., the overall daylighting potential of the building, the occurrence of excessive illuminances or luminances, as inputs to behavioral models for light switching and/or blinds usage, and the potential of daylight responsive lighting controls to reduce building energy usage. Thus a daylight performance metric would need to be based on a time-series of instantaneously occurring daylight illuminances since these cannot be reliably inferred from cumulative values. As noted, evaluations should

span an entire year. There is some debate as to whether the daily time period of analysis should be all daylit hours, which vary in length with the seasons, a standardized "working day" of 8, 10, or 12 h, or the actual occupancy pattern of the space. Different purposes are likely to favor different daily analysis periods.

How climate-based modeling compares to the standard method (daylight factors) and the more recent "clear sky options" is shown in Fig. 20. A daylight factor value can be converted into an estimate of the overall occurrence of daylight in a space by using a diffuse illuminance availability curve for that locale (as described in an earlier section). However, since the contribution of sunlight cannot be accounted for, the method must be considered a crude estimator. The various "clear sky" options generally do not require that the skies are normalized. Thus there is no link with prevailing climate. In contrast, climate-based daylight modeling uses the full year of irradiance or illuminance data found in the climate file, and derives the instantaneous sky and sun conditions from these.

Daylight Metrics

The purpose of a metric is to combine various factors that will successfully predict better or worse performance outcomes and so inform decision making.



Daylight, Indoor Illumination, and Human Behavior. Figure 20
Climate-based daylight modeling compared to daylight factor and snap-shot "clear sky" options. Luminous efficacy model* is used if the data contains irradiance only

Performance may be described by more than one metric, i.e., it is not necessary to combine all significant factors into one metric. This implies a preference for simplicity so they can be intuitively understood, and a direct relation to measurable outcomes made. When metrics are sufficiently refined and understood and their predictive capabilities validated, then performance criteria can be set for various guidelines and recommendations.

As has been noted, metrics founded on the daylight factor are relatively straightforward since there is no time-varying component and so they simply report on the DF value at a point, some average DF value across a workplane, or perhaps some measure of uniformity of the DF across the workplane. Metrics founded on climate-based modeling are potentially far more complex since the simulations output illuminance data at each time-step for every point in the space. Thus, for all daylight hours in the year, a climate-based simulation would output approximately 4,380 values for every calculation point considered, and potentially several times this number if the simulations were run at a shorter time-step to, say, better resolve the progression of the solar patch across the internal space.

Various climate-based daylight metrics have been formulated since the emergence of climate-based modeling in the late 1990s. As of 2010, these metrics are being investigated by daylighting researchers in order to determine their potential to reliably characterize daylight in buildings for the purpose of discriminating between "good," "bad," and "mediocre" designs [61, 68]. One of the more straightforward climate-based metrics is daylight autonomy (DA) [69]. The DA metric determines the annual occurrence (within, say, working hours) of illuminances above a stated design level illuminance, e.g., 300 or 500 lux. It is well known, however, that occupants prefer daylight illumination not to exceed certain levels, although it is not clear what precisely those levels are since occupants vary greatly in their responses. The "useful daylight illuminance" (UDI) metric was formulated as a means to reduce the voluminous time-series data from a climate-based simulation to a form that is of comparative interpretative simplicity to the daylight factor method, but which nevertheless preserves a great deal of the significant information content of the illuminance time-series. The UDI metric informs on the occurrence of illuminances in the range that occupants either prefer or tolerate together with the propensity for excessive levels of daylight that are associated with occupant discomfort and unwanted solar gain [62]. Thus useful daylight illuminance is more firmly grounded on human factors than metrics which determine only sufficiency for task.

Achieved UDI is defined as the annual occurrence of illuminances across the workplane that are within a range considered "useful" by occupants. The range considered "useful" is based on a survey of reports of occupant preferences and behavior in daylit offices with user-operated shading devices. Daylight illuminances in the range 100–500 lux are considered effective either as the sole source of illumination or in conjunction with artificial lighting. Daylight illuminances in the range 500 to around 2,000 or maybe 3,000 lux are often perceived either as desirable or at least tolerable.

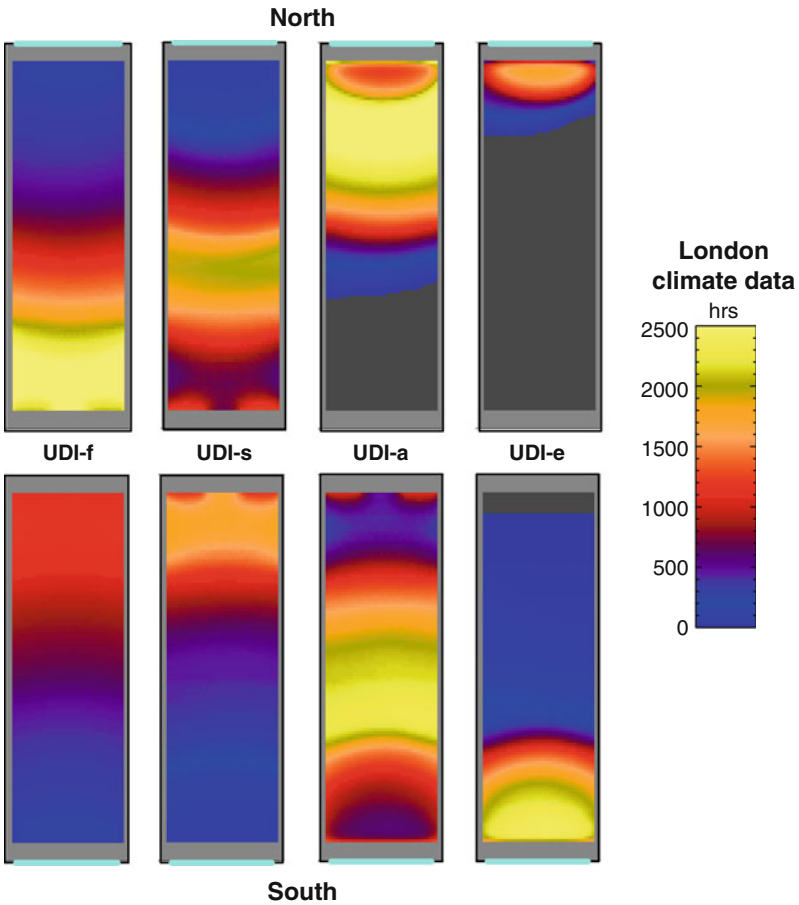
The range limits for UDI depend to a degree on the particular application, and, at the time of writing, there is no consensus on what precise values the upper and lower range limits should have. Nonetheless, the UDI scheme combines intuitive simplicity with rich information content. For the example shown below, UDI was defined as the annual occurrence of daylight illuminances that are between 100 and 3,000 lux. The UDI range is further subdivided into two ranges called UDI-supplementary and UDI-autonomous. UDI-supplementary gives the occurrence of daylight illuminances in the range 100–300 lux. For these levels of illuminance, additional artificial lighting *may* be needed to supplement the daylight for common tasks such as reading. UDI-autonomous gives the occurrence of daylight illuminances in the range 300–3,000 lux where additional artificial lighting will most likely not be needed. The UDI scheme is applied by determining at each calculation point the occurrence of daylight levels where:

- The illuminance is less than 100 lux, i.e., UDI "fell-short" (or UDI-f)
- The illuminance is greater than 100 lux and less than 300 lux, i.e., UDI supplementary (or UDI-s)
- The illuminance is greater than 300 lux and less than 3,000 lux, i.e., UDI autonomous (or UDI-a)
- The illuminance is greater than 3,000 lux, i.e., UDI exceeded (or UDI-e)

The same space that was used for the daylight factor example is employed again. Now, however, time-varying illuminances across the workplane during the hours of 09:00 to 18:00 were predicted for the entire year, i.e., a total of 3,285 h. Simulations were carried out for the space having, in turn, north- and south-facing glazing and the sun and sky conditions were derived from a standardized climate file for London, UK. For this illustration, there was no attempt to model user operation of blinds, etc.

The four UDI metrics determined from the illuminance data for both orientations of the space are shown in Fig. 21. The occurrence of the various UDI metrics is shown using color, and a zero value for the metric is shaded dark gray. The light gray perimeter indicates the space between the work and the walls of the space. In these plots we can readily see the contribution that sunlight has to overall illumination of the space by comparing the north and south orientations. In the absence of target values for the various UDI metrics, the approach is most useful to the designer when, e.g., various facade options need to be compared. Those that maximize the UDI-a metric without undue occurrence of UDI-e are likely to offer the best daylighting design – though it needs to be acknowledged that this supposition will need to be tested and proven.

Although there is no consensus yet on the type of climate-based daylight metric (e.g., UDI and DA) that should feature in future guidelines [68], there is



Daylight, Indoor Illumination, and Human Behavior. Figure 21
Annual occurrence of UDI metrics between the hours of 09:00 and 18:00 for north and south orientations (London, UK, climate data)

a groundswell of opinion that the half-century-old basis of daylight evaluation needs to be updated. In the meantime, climate-based modeling has made the transition from research to practice and is increasingly used by consulting engineers in their striving to achieve that elusive goal of the well-tempered daylit environment.

Metrics founded on climate-based daylight modeling will be the key to reliable evaluation of designs for well-daylit low-energy buildings. There remains, however, some work to ensure that future daylighting metrics will act in concert with other metrics, e.g., for thermal performance, and so not give conflicting guidance to the building designer.

Advanced Glazing Systems and Materials

The use of daylight in office buildings is generally considered to be a greatly underexploited resource. In large part this is because of both the highly variable nature of daylight illumination and its prevailing direction as it enters a space. Variability in daylight means that users will often need to use shades at least some of the time to moderate excessive ingress of daylight. Daylight illumination from perimeter glazing enters a space with a predominant downward direction. Those work areas close to the window may receive plentiful daylight; however, much of the natural light will arrive at the floor where it will be mostly absorbed due to the typically low reflectances of flooring materials. What little light there is reflected up from the floor may encounter significant obstructions (e.g., desk, chairs) before having the possibility of being reflected again – now downward off the ceiling – to contribute to illumination on the workplane. In short, reflection of light from the floor to help illuminate the space is a low efficiency process with limited effectiveness. Another key issue with traditional building facades that also serves to greatly impair the potential daylighting performance of a space is the design and user operation of the shading systems, e.g., venetian blinds, etc. Many shading systems act as a "shutter" that is either open or closed, with users rarely making the effort to optimize the shading for both daylight provision and solar/glare control. Furthermore, blinds are often left closed long after the external condition has changed. The exploitation of daylight in buildings, particularly those dominated by side-lighting from vertical windows, could be greatly improved by any of the following:

1. Redirecting the daylight that enters the space toward the ceiling and walls where subsequent reflections will help to better distribute the daylight deeper into a space
2. For a shading system, modulating the daylight gradually rather than an on/off shutter operation
3. Reducing, possibly even eliminating, the need for user interventions, e.g., the lowering of blinds

The purpose of the majority of so-called advanced glazing systems or materials (AGSM) is to improve overall daylighting in a space by achieving one or more of the above-mentioned goals.

AGSMs fall into to two broad categories: active and passive. Active systems vary some property (e.g., visible light transmittance) automatically according to some control parameter (e.g., illumination at the workplane), though they may often include an option for user override. A passive system is one that is invariably fixed requiring no external control either automatic or manual. However, some of the passive systems described below could have, say, automated movement though this would add considerably to the cost. Any of the systems discussed below will be subject to the same issues regarding visual quality (i.e., providing adequate views, avoiding glare, etc.) that are a consideration for conventional glazing.

Passive AGSMs The goal of most of the passive systems is to redirect daylight in some fashion, and any changes in the magnitude and distribution of the transmitted light is due to variation in the amount and direction of the incoming light. Passive AGSMs are usually in the form of a material applied to one of the glass surfaces of a window, or are themselves a glazing element, e.g., one part of a double glazing system where the other might be standard glass. The light redirection can be achieved by manipulating the specular and/or diffuse transmission properties of the material. Glazing systems in this category include light redirecting prismatic panels, laser-cut panels, diffusing materials, mirrored louvers, etc. Some of these materials have been available since the 1980s; however, the uptake has not been great. In part this is due to either limited

long-term performance data [70] and/or the inability to predict their daylighting performance at the design stage due to insufficient knowledge of the material's light transmission properties [71].

Anecdotal reports from some early installations of prismatic glazing suggests that the lack of a clear view to the outside together with secondary effects such as dispersion have not found favor with building occupants. More recent innovations have attempted to address these issues. The specular redirecting material marketed as Serraglaze provides a relatively clear view for an observer at normal incidence, but effectively blocks direct transmission of high-angle sun by redirecting it up toward the ceiling. Translucent materials such as Kalwall have very low thermal conductivity and can be used to replace sections of wall as well as glazing. However, these materials are essentially diffusing panels and so must be used in conjunction with clear glazing to provide views of the outside.

A common addition to windows to moderate both the solar gain and the daylight ingress is a film to reduce the overall transmissivity of the glazing. These films vary greatly in visible transmittance (from around 0.6 down to 0.1) and may have a pronounced color hue. Even films with the lowest transmittance usually need additional shading to moderate direct sun. The reduced transmittance of windows with films can, for the occupants, make views to the outside appear drab and gloomy, particularly on overcast days. A new approach to tempering daylight which, like films, can be applied as a retrofit is a novel treatment called Solaveil [72]. This material is fabricated using digital printing techniques which deposit microscale 3D structures on the substrate which act both as a "glare filter" and a redirecting "micro-light shelf." Before and after photographs for a retrofit installation of Solaveil are given in Fig. 22. The "before" image on the left shows how the building was typically used prior to the retrofit: the blinds are down to control glare and direct sun, and the electric lights are switched on. The "after" image on the right shows the treated area of the window which now acts as a diffusing light shelf redirecting light to the ceiling and protecting occupants from harsh, direct sun. Note that no additional shading has been installed and the lights – now photoelectrically controlled – are switched off. The lower untreated part of the window provides occupants with views to the outside. Initial findings indicate a significant potential for saving energy in both lighting and cooling from such interventions.

Active AGSMs The most established of the technologies in this category is the automated shading system where, say a motorized roller blind is deployed incrementally according to some sensor input, e.g., measured daylight level. This shading system features in the facade design throughout the majority of the 52 floors of the New York Times Building [73]. The design goals for the shading system were to

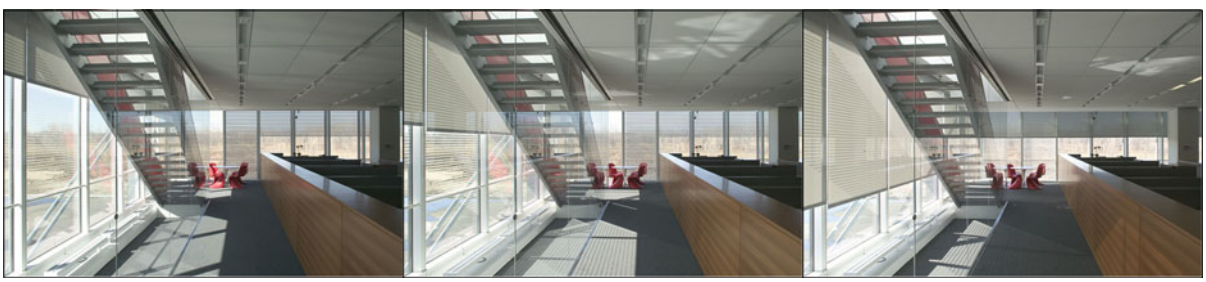
Daylight, Indoor Illumination, and Human Behavior. Figure 22
Photographs showing before and after cases where the standard glazing with blinds was replaced with a novel window treatment called Solaveil (Photos courtesy Solaveil)

- Maximize natural light
- Maximize occupant connectivity with the outdoors, i.e., external views
- Intercept sunlight penetration so as to avoid direct solar radiation on the occupants
- Maintain a glare-free environment
- Provide occupant manual override capability

The overall intent was to keep the shades up as much of the time as possible without causing thermal or visual discomfort. Thermal comfort is assured by solar tracking and the geometry of the external sun screens. Visual comfort is attained by managing the luminance on the window wall so that it does not exceed certain threshold values. A manual override system was specified because previous post occupancy evaluations of automated shade systems indicated that occupants were likely to complain if a manual override was not provided. A sequence of photographs showing incremental deployment of the shades is given in Fig. 23. Although a formal post-occupancy evaluation of the New York Times Building has yet to be carried out, anecdotal evidence from informal surveys indicates a high level of user satisfaction with the daylighting systems. Furthermore, effective daylighting has significantly reduced the energy consumed for artificial lighting. The NYT daylighting system provides a degree of modulation for the shading (they are deployed by increments) and has greatly reduced the need for user interventions.

In contrast to using standard shade materials deployed either manually or automatically, a glazing with a transmissivity that varies continuously between clear and dark extremes would offer a much greater degree of control over the luminous environment. Indeed, the dynamic control of daylight has been termed the "Holy Grail of the fenestration industry" [74]. In principle, the approach is simple: the transmission properties of the glazing are varied to achieve the best possible luminous environment. Formulations based on electrochromic (EC) principles, where the glazing transmission is modulated by a small applied voltage, are considered the most promising at present. In practice, the formulation and production of commercial-sized EC glazing has proved a formidable task. Recently, however, a number of technical hurdles have been overcome, preproduction samples of EC glazing have been deployed in test facilities for evaluation [75] and commercial installations have followed, Fig. 24. The optical properties of EC windows can be modulated using control variables such as incident or transmitted solar radiation, daylight illuminance, ambient air temperature, or space thermal load [76]. In the clear state the visible transmittance of EC glazing can be as high as 60%. However, to avoid the need for any additional shading (e.g., by venetian blinds), the visible transmittance in its tinted state has to be as low as 2%. This has now been achieved in production samples and, as prices fall, it is to be expected that EC glazings will become a more mainstream product.

In addition to the electrochromic method, there are also formulations for light modulating glazing where the visible transmittance varies autonomously in response to either the temperature of the material or the intensity of incident illumination, known as thermotropic and phototropic respectively. The phototropic glazings have a formulation which is similar to that commonly used in "reactive" sunglasses, whereas thermotropic glass consists of two panes of glass sandwiching a polymer gel that undergoes a transition from clear to cloudy at a threshold



Daylight, Indoor Illumination, and Human Behavior. Figure 23
Sequence showing deployment of shades in a full-size mock-up of New York Times Building offices (Photos courtesy LBNL)

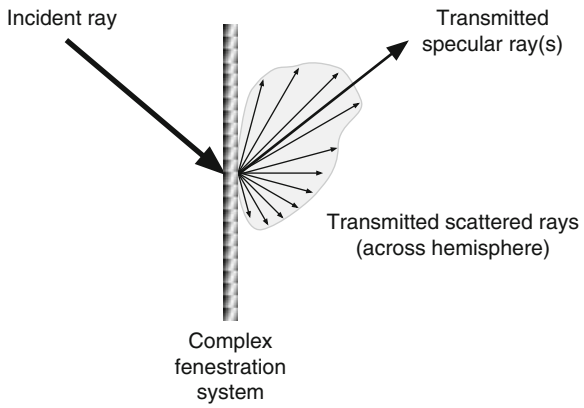
Daylight, Indoor Illumination, and Human Behavior. Figure 24
Images showing electrochromic glazing in clear and darkened state (Photos courtesy SAGE Electrochromics)

temperature [77]. These autonomous systems are largely confined to research studies at present, and the inability to control their transmission properties depending on some internally measured quantity, e.g., daylight levels, may ultimately prove to be a drawback rather than an advantage.

Evaluating AGSMs Advanced glazing systems/materials have the potential to enhance the daylighting performance of a space. Performance, cost, and user acceptance are key factors which determine the overall effectiveness of an installation. The first two of these factors can, in principle, be determined using simulation – the predicted performance could be judged against likely cost. It may even be possible to employ simulation to anticipate to some degree user acceptance of a novel glazing system/material, though the evaluation would have to be founded data from occupant-based studies to have credibility.

To simulate the performance of an advanced glazing system/material can be quite challenging due to their particular optical properties which often diverge greatly from standard glazing. The optical properties of ordinary clear glazing and reflective materials that have a matt finish are relatively easy to characterize for the purpose of lighting simulation. Less straightforward materials such as coated glazings and materials that produce part specular reflections are more challenging to both characterize and also to model accurately in a simulation. Tools such as Optics 5 and Window 6 can assist the creation of the necessary material description files for multilayer coated glazings [78]. The highlights resulting from even tiny specular reflections are an important part of the overall visual impression of a daylit space; however, the total light energy resulting from these reflections is usually very small and can be ignored when predicting illuminance quantities. Specular reflections are only important for overall light transfer in a space when significant amounts of the entrant direct and diffuse light are reflected, e.g., when a mirror light shelf is present. Large-scale reflecting/redirecting features such as light shelves or "skylight" wells can be modeled using standard *Radiance*.

A major issue with advanced glazing systems/materials however is that there is usually no straightforward relation between incident and transmitted light that can be determined a priori from simple, e.g., analytical, methods. Thus the optical properties of the AGSM need to be determined from either comprehensive measurements or, alternatively, simulation. For each light ray incident on an AGSM, there may be one or more strongly transmitted rays – which may be redirected in some fashion – together with, in most cases, a unique distribution of semi-diffuse or scattered light, Fig. 25. Thus, to fully characterize the material, the distribution in luminous output across the full hemisphere of transmitted rays needs to be determined for every incident direction [79]. This is the bidirectional transmission distribution function or BTDF [80]. The BTDF is challenging to characterize even for seemingly simple materials such as translucent



Daylight, Indoor Illumination, and Human Behavior. Figure 25
Schematic showing the distribution in transmitted light from a complex fenestration system

glazing [81]. Another approach to characterization of the BTDF is to predict it by simulation rather than by measuring it directly [82]. For this approach, the geometric microstructure of the material needs to be specified to a high degree of precision and the BTDF predicted using a forward ray-tracing program.

It is possible to model some AGSMs without going to the lengths of determining the full BTDF if the transmission properties of the materials can be adequately represented by an analytical function. This has been achieved for certain types of laser cut panels [83] and the redirecting material Serraglaze [84]. A limited set of angular-dependent transmission measurements will still need to be taken to calibrate the analytical model.

Note that although there are various approximations to model the light transmission through venetian blinds, even these commonplace devices have complex optical properties. Both the slat angle and the coverage of venetian blinds can be varied continuously and independently of each other. For any given sun position, either of these factors has a considerable effect on the overall light transmission, i.e., the BTDF [85]. Venetian blinds therefore can be more difficult to model accurately than many of the "advanced" systems/materials because their BTDF is dependent on the user operation.

Light-pipes (i.e., tubular daylighting devices) offer a potentially effective daylighting strategy for low-rise buildings. The performance of a light-pipe can be estimated using analytical methods or relatively simple software tools [86, 87]. The detailed simulation of light-pipe performance however remains quite challenging.

Characterization of BTDFs by either measurement or prediction is a highly specialized task, as is the use of these complex transmittance data in lighting simulations. There is considerable research to be carried out at all stages from characterization to implementation in a software tool before their use in lighting simulation becomes commonplace. The development of libraries of BTDF databases for various products, based on standardized test procedures, will be necessary to enable full utilization of these products in design optimization studies.

New Approaches to Measuring the Daylit Environment

Quantitative knowledge of the internal daylit environment has, until recently, been largely restricted to measurements of illuminance (typically at the workplane) and "spot" measurements of luminance taken by a narrow (e.g., 1°) field-of-view photometer. Illuminance at the workplane is rarely recorded over long periods in an occupied building due to the cost of the monitoring and the disruption caused by wires, etc. (though there are now largely autonomous wireless sensors available, albeit at a high cost). Humans have an almost 180° forward-facing field of view, so a measurement taken using a "spot" photometer with a 1° field-of-view records only a very small part of what is seen.

As noted in an earlier section, understanding of visual comfort in daylit environments is lacking in part because we have scant empirical data on the (constantly changing) visual field. A recent technology called high dynamic range (HDR) imaging has greatly expanded our capacity to measure and describe the visual field. An HDR image is one where every pixel contains a luminance reading for that point in the recorded scene, in other words: a measurement of luminance. There are a small number of specialist HDR cameras on the market; however, it is possible to create HDR images from multiple exposures taken by consumer digital cameras which can have up to

10 million or more pixels [88]. Furthermore, the consumer cameras can be fitted with a full fish-eye lens so the recorded image will be equivalent to or exceed the human field of view.

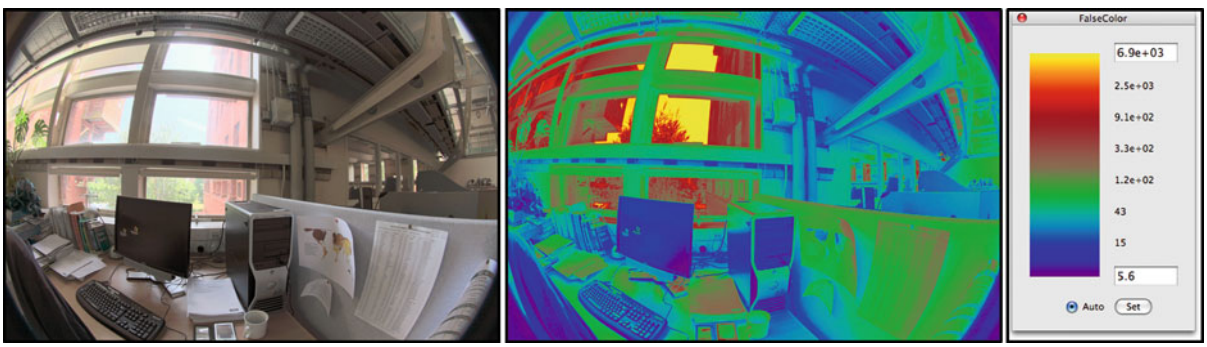
HDR image capture has been used in a number of daylight glare studies [89, 90] and also more generally to investigate lighting preferences in office spaces [91]. An example HDR image alongside a standard digital photograph is shown in Fig. 26. The view is of a workstation close to a window and with direct view of the sky. From the false-color HDR image it can be seen that the luminance of the sky is of the order of 7,000 cd m^{-2}. HDR imaging techniques were used to calibrate and commission the dynamic shading systems of the New York Times Building [73]. A key consideration in the design of the control system for the automated shades was to allow in as much daylight as possible without exceeding visual comfort criteria based on field-of-view (i.e., perceived) luminance. This was tested by measurement in the full-size mock-up of offices (Fig. 23) and evaluated more generally for the building using simulation (i.e., climate-based daylight modeling). The mock-up allowed only for limited scenarios, i.e., just two view directions with fixed external obstruction. The simulation, however, allowed multiple floors of the building to be evaluated in context (i.e., the surrounding buildings) and for all possible view directions. The images in Fig. 27 show long, medium, and close-up views of the highly detailed 3D model used for the simulation – model detail of the offices was generated on a per-floor basis for each of the floors evaluated.

High dynamic range imaging technology may become sufficiently compact and inexpensive to replace the traditional sensors that are currently used with daylight-responsive systems, e.g., photocell-controlled artificial lighting, automated shades/blinds and the transmission of electrochromic glazing [92].

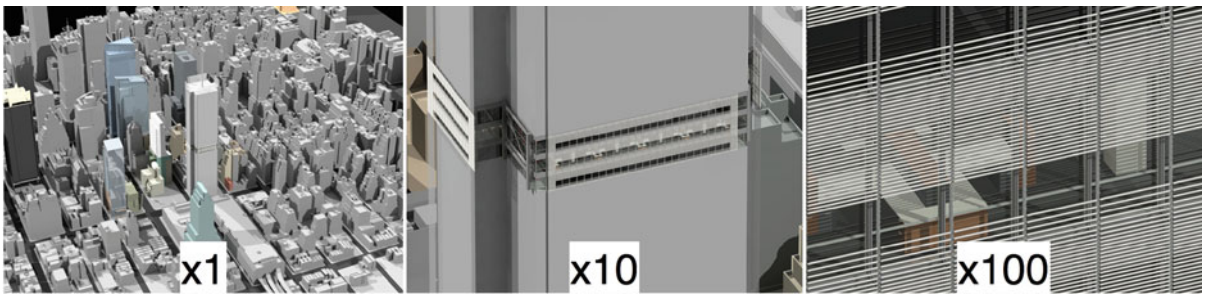
Daylight and Saving Energy

It seems to be generally believed that "good" daylighting design will lead to reductions in electric lighting consumption, and also overall energy consumption. This belief results in part from common-sense notions and the pioneering work of Crisp and Hunt in the 1970s [93, 94]. The potential for energy savings was usually based on extrapolating internal illuminance from daylight factors and cumulative daylight distributions, and then applying some model of lighting control [95]. Lighting control models based on manual switching were derived from observed patterns of behavior [57, 96]. It was realized early on that occupant control alone was unlikely to lead to significant energy savings for the simple reason that lights were likely to remain switched on even when there was plentiful daylight. Some form of timed switching and/or automatic control was needed to ensure energy savings, and a number of largely theoretical formulations for occupancy-sensor and photoelectric control of lighting were devised [93]. The design and artificial lighting of nondomestic spaces has changed considerably over the last 30 years, and some of the findings noted in occupancy studies carried out three or more decades ago may not necessarily hold today.

Post-occupancy studies carried out in real buildings have shown that the actual energy performance is



Daylight, Indoor Illumination, and Human Behavior. Figure 26
High dynamic range image used to measure the luminance of the occupants field of view



Daylight, Indoor Illumination, and Human Behavior. Figure 27
Highly detailed 3D model of the New York Times Building and surrounding Manhattan – full office detail modeled for floor 26

invariably markedly worse than that predicted at the design stage. The landmark PROBE study determined many of the reasons for this [97]. Some of the findings specific to lighting controls are noted below:

- Default states which are non-optimal, but cause the least trouble for occupants and management. The most common of these is blinds closed lights on, which has undermined many a daylight and lighting control strategy.

  Photocells used for perimeter dimming ... were also confused by light redirected upwards onto them from the venetian blind slats, requiring control setpoints to be raised, so reducing the benefits of daylight-linked dimming.

The ratio of predicted to realized energy savings is defined as the "realized savings ratio" or RSR. Studies of automatic photocontrol performance in the USA have shown very high RSRs for simple top-lit spaces, and much lower RSRs for more complex side-lit spaces [98]. Predicting the performance of an automated lighting control system is a function of many factors, including not only space design and daylight availability, but also lighting system design, control settings, commissioning history, and occupant override behavior.

Daylight is merely the visible part of the radiant energy that enters through windows. Furthermore, the bulk of the daylight energy that enters a space is converted into thermal energy after just a few reflections. Many office buildings in moderate climates now have air-conditioning largely due to the high internal gains. In warmer climates cooling may be needed for large parts of the year. When cooling is needed in a space, both the use of electric lighting and the ingress of daylight will each add to the cooling load.

Attempts to provide good daylighting could therefore lead to a net *increase* in energy consumption if the additional cooling load due to daylight (i.e., including the solar component) exceeds the energy saved due to reduced electric lighting, or if the net heat gains and losses through the fenestration do not compensate for the lighting energy saved. In fact, an all too common scenario in over-glazed buildings is where the blinds are down to control glare and the lights are left on. This leads to the undesirable combination of high solar gains (blinds reject only a small part of the energy once it has passed through the glazing) and no "daylight benefit" in terms of displaced lighting energy or daylight provision. A full consideration of the potential for daylighting to save energy should, at some point, account also for the thermal effects of daylight. In which case, daylight metrics will need to be calibrated against criteria for whole building energy use and not just the potential to reduce the energy consumed for electric lighting. That was not a possibility with the standard daylight factor approach because the sun did not figure in the evaluation. However, with the emergence of climate-based daylight modeling, a truly holistic evaluative schema which can be applied at the design stage is now a likely prospect.

Good daylighting alone is unlikely to save energy unless it is part of an integrated design scheme. The typical lighting power densities (LPDs) in office spaces range from about 12–20 W/m^2, with those at the lower end considered "good practice." It is possible however

to achieve LPDs significantly lower than the good practice value – without recourse to emerging technologies such as light emitting diodes – using only good-quality low-energy fluorescent lights. This was successfully demonstrated in the New York Times Building which has an LPD of only 4.26 W/m^2.

A fully integrated low-energy daylighting design must necessarily be tailored to the local environment, in terms of both the climate and the surrounding built context. It is hoped that a new generation of daylight metrics founded on climate-based modeling will help designers to achieve that elusive balance between daylight provision and effective solar protection. And, importantly, offer design guidance that does not conflict with other criteria such as thermal performance of the building.

The potential for new technologies (e.g., electrochromic glazing) to save energy can only be reliably assessed using climate-based daylight modeling. Building facades may become electricity generators through the widespread adoption of various building-integrated photovoltaic (PV) technologies [99]. Semitransparent (PV) panels could serve as combined widows and electricity generators [100]. Thus daylight consideration could become closely linked with other performance aspects of the overall building.

Future Directions

Both the theoretical basis for and the practical application of daylighting in buildings are in the midst of a fundamental reappraisal. The emergence of findings that relate daylight exposure to positive outcomes in terms of productivity, health, and well-being have led to a renewed focus on daylighting in buildings. Throughout the developed world, governments and regulatory bodies are encouraging the adoption of design guides and recommendations in the hope that reduced energy consumption in buildings can be achieved, in part, through effective daylighting design. The theoretical basis for these guidelines, however, is a crude representation of actually occurring daylight conditions. A more realistic evaluative scheme has emerged with the development of climate-based daylight modeling, which it is widely believed will soon replace the daylight factor as the basis of a new generation of daylighting guidelines. A large number of advanced glazing systems and materials for enhancing the daylight in buildings have recently appeared on the market, and there are new formulations in the early stages of development. Market penetration of innovative daylighting systems has, until now, proven to be difficult because the standard "measure" of performance (i.e., the daylight factor) gives no indication of how much natural light and how often. Data on the magnitude and occurrence of absolute measures of natural illumination – precisely how much and how often – are vital to reliably assess both the performance-effectiveness and the cost-effectiveness of daylighting systems. Thus the hoped-for emergence of climate-based daylight metrics will greatly assist in the evaluation of these daylighting systems, and, for those shown to be effective, their marketing. Another consideration is the advances in measurement and control through techniques such as high dynamic range imaging. These theoretical and technological advances have the potential to radically improve our perception of what constitutes good daylighting in buildings, both in terms of basic design parameters and the use of novel glazing materials, thus paving the way for daylighting guidelines and codes that lead to the reliable and robust production of truly healthy, low-energy buildings.

Bibliography

1. Baker NV, Fanchiotti A, Steemers KA (1993) Daylighting in architecture: a European reference book. James & James, London. ISBN 1-873936-21-4
2. Mardaljevic J, Andersen M, Roy N, Christoffersen J (2011) Daylighting metrics for residential buildings. In: CIE 27th session, Sun City, South Africa
3. Al Marwaee M, Carter DJ (2006) A field study of tubular daylight guidance installations. Light Res Technol 38(3):241–258
4. CIBSE Guide A (2006) Environmental design. Chartered Institution of Building Services Engineers, London
5. Hopkinson RG (1963) Architectural physics – lighting. Her Majesty's Stationery Office, London
6. Moon P, Spencer DE (1942) Illuminations from a non-uniform sky. Illum Eng 37:707–726
7. Enarun D, Littlefair P (1995) Luminance models for overcast skies: assessment using measured data. Light Res Technol 27(1):53–58
8. Dufton AF, Beckett HE (1932) The heliodon: an instrument for demonstrating the apparent motion of the sun. J Sci Instrum 9(8):251
9. Ward Larson G, Shakespeare R (1998) Rendering with radiance: the art and science of lighting visualization. Morgan Kaufmann, San Francisco

10. Mardaljevic J (2004) Verification of program accuracy for illuminance modelling: assumptions, methodology and an examination of conflicting findings. Light Res Technol 36(3):217–239
11. Harris L (2007) Anstey's rights of light and how to deal with them, 4th edn. RICS Books, London
12. Lynes JA (1979) A sequence for daylighting design. Light Res Technol 11(2):102–106
13. Crisp VHC, Littlefair PJ (1984) Average daylight factor prediction. In: Proceedings of national lighting conference, Cambridge. CIBSE, London
14. Reinhart CF, LoVerso VRM (2010) A rules of thumb-based design sequence for diffuse daylight. Light Res Technol 42(1):7–31
15. Cannon-Brookes SWA (1997) Simple scale models for daylighting design: Analysis of sources of error in illuminance predictiont. Light Res Technol 29(3):135–142
16. Mardaljevic J (2002) Quantification of parallax errors in sky simulator domes for clear sky conditions. Light Res Technol 34(4):313–327
17. Mardaljevic J (1995) Validation of a lighting simulation program under real sky conditions. Light Res Technol 27(4): 181–188
18. Mardaljevic J (2000) Daylight simulation: validation, sky models and daylight coefficients. PhD thesis, De Montfort University, Leicester
19. Mardaljevic J (2001) The BRE-IDMP dataset: a new benchmark for the validation of illuminance prediction techniques. Light Res Technol 33(2):117–134
20. Reinhart CF, Fitz A (2006) Findings from a survey on the current use of daylight simulations in building design. Energ Buildings 38(7):824–835
21. Ibarra D, Reinhart CF (2009) Daylight factor simulations – how close do simulation beginners 'really' get? In: International Building Performance Simulation Association conference, Glasgow, Scotland, July 27–30, pp 196–203
22. BSI (2008) Lighting for buildings Code of practice for daylighting BS 8206–2. British Standards Institute, London
23. Hunt DRG (1979) Availability of daylight. BRE Report BR21, Building Research Establishment
24. Building Research Establishment (1985) Lighting controls and daylight use. BRE Digest 272
25. Building Research Establishment. BREEAM – the BRE environmental assessment method. http://www.breeam.org. Accessed 2011
26. The U.S. Green Building Council. LEED: leadership in energy and environmental design. http://www.usgbc.org/LEED. Accessed 2011
27. Littlefair PJ (1998) Predicting lighting energy use under daylight linked lighting controls. Build Res Inf 26(4):208–222
28. Clarke JA (2001) Energy simulation in building design, 2nd edn. Butterworth-Heinemann, Oxford
29. Littlefair PJ (1985) The luminous efficacy of daylight: a review. Light Res Technol 17(4):162–182
30. Skartveit A, Olseth JA (1992) The probability density and autocorrelation of short-term global and beam irradiance. Solar Energy 49(6):477–487
31. Begemann SHA, den Beld GJ, Tenner AD (1997) Daylight, artificial light, and people in an office environment: Overview of visual and biological responses. Int J Ind Ergon 20(3): 231–239
32. Galasiu AD, Veitch JA (2006) Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review. Energ Buildings 38(7): 728–742
33. LG7 CIBSE/SLL (2005) Lighting guide 7: office lighting. Chartered Institution of Building Services Engineers, London
34. EWCO (2005) Annual review of working conditions in the EU: 2004–2005. Technical report, European Foundation for the Improvement of Living and Working Conditions
35. CIBSE (1996) Lighting Guide 3: the visual environment for display screen use, addendum 2001. Chartered Institution of Building Services Engineers, London
36. CIBSE/SLL (2002) Code for lighting. Chartered Institution of Building Services Engineers/Society of Light and Lighting, London
37. Osterhaus WKE (2005) Discomfort glare assessment and prevention for daylight applications in office environments. Solar Energy 79(2):140–158
38. Chauvel P, Collins JB, Dogniaux R, Longmore J (1982) Glare from windows: current views of the problem. Light Res Technol 14(1):31–46
39. Velds M (2002) User acceptance studies to evaluate discomfort glare in daylit rooms. Solar Energy 73(2):95–103
40. Tuaycharoen N, Tregenza PR (2007) View and discomfort glare from windows. Light Res Technol 39(2):185–200
41. Webb AR (2006) Considerations for lighting in the built environment: Non-visual effects of light. Energ Buildings 38(7): 721–727
42. Collins BL (1976) Review of the psychological reaction to windows. Light Res Technol 8(2):80–88
43. Boyce PR, Beckstead JW, Eklund NH, Strobel RW, Rea MS (1997) Lighting the graveyard shift: The influence of a daylight-simulating skylight on the task performance and mood of night-shift workers. Light Res Technol 29(3):105–134
44. Brainard GC, Hanifin JP, Greeson JM, Byrne B, Glickman G, Gerner E, Rollag MD (2001) Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. J Neurosci 21(16):6405–6412
45. McIntyre IM, Norman TR, Burrows GD, Armstrong SM (1989) Human melatonin suppression by light is intensity dependent. J Pineal Res 6(2):149–156
46. Rea MS, Figueiro MG, Bullough JD (2002) Circadian photobiology: an emerging framework for lighting practice and research. Light Res Technol 34(3):177–187
47. Pechacek CS, Andersen M, Lockley SW (2008) Preliminary method for prospective analysis of the circadian efficacy of (day) light with applications to healthcare architecture. Leukos 5(1):1–26
48. Hubalek S, Brink M, Schierz C (2010) Office workers' daily exposure to light and its influence on sleep quality and mood. Light Res Technol 42(1):33–50

49. Walch JM, Rabin BS, Day R, Williams JN, Choi K, Kang JD (2005) The effect of sunlight on postoperative analgesic medication use: a prospective study of patients undergoing spinal surgery. Psychosom Med 67(1):156–163
50. Commission for Architecture and the Built Environment (2004) The role of hospital design in the recruitment, retention and performance of NHS nurses in England. CABE, pp 120–124
51. Heschong Mahone Group (1999) Daylighting in schools: an investigation into the relationship between daylight and human performance. Pacific Gas and Electric Company, San Francisco
52. Heschong Mahone Group (2003) Windows and offices: a study of office worker performance and the indoor environment. Heschong L (ed) CEC, California Energy Commission
53. Heschong Mahone Group (1999) Skylight and retails sales, an investigation into relationship between daylighting and human performance. Pacific Gas and Electric Company, San Francisco
54. National Research Council (2007) Green schools: attributes for health and learning. National Academies Press, Washington, DC, Committee to Review and Assess the Health and Productivity Benefits of Green Schools
55. Noguchi H, Itou T, Katayama S, Koyama E, Morita T, Sato M (2004) Effects of bright light exposure in the office (proceedings of the 7th international congress of physiological anthropology). J Physiol Anthropol Appl Hum Sci 23(6):368
56. Mills P, Tomkins S, Schlangen L (2007) The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. J Circadian Rhythms 5(1):2
57. Hunt DRG (1980) Predicting artificial lighting use – a method based upon observed patterns of behaviour. Light Res Technol 12(1):7–14
58. Reinhart CF, Voss K (2003) Monitoring manual control of electric lighting and blinds. Light Res Technol 35(3):243–258
59. Newsham GR, Mahdavi A, Beausoleil-Morrison I (1995) Lightswitch: a stochastic model for predicting office lighting energy consumption. In: Proceedings of right light 3, Newcastle-upon-Tyne, UK
60. Reinhart CF (2004) Lightswitch-2002: a model for manual and automated control of electric lighting and blinds. Solar Energy 77(1):15–28
61. Mardaljevic J, Heschong L, Lee E (2009) Daylight metrics and energy savings. Light Res Technol 41(3):261–283
62. Mardaljevic J (2006) Examples of climate-based daylight modelling. In: CIBSE national conference 2006: engineering the future, Oval Cricket Ground, London, 21–22 Mar 2006
63. Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, Strand RK, Liesen RJ, Fisher DE, Witte MJ, Glazer J (2001) EnergyPlus: creating a new-generation building energy simulation program. Energ Buildings 33(4):319–331
64. Mardaljevic J (2006) Time to see the light. Build Serv J September:59–62
65. Mardaljevic J (2000) Simulation of annual daylighting profiles for internal illuminance. Light Res Technol 32(3):111–118
66. Reinhart CF, Herkel S (2000) The simulation of annual daylight illuminance distributions – a state-of-the-art comparison of six RADIANCE-based methods. Energ Buildings 32(2):167–187
67. Nabil A, Mardaljevic J (2006) Useful daylight illuminances: a replacement for daylight factors. Energ Buildings 38(7):905–913
68. Reinhart CF, Mardaljevic J, Rogers Z (2006) Dynamic daylight performance metrics for sustainable building design. Leukos 3(1):7–31
69. Reinhart CF, Walkenhorst O (2001) Validation of dynamic radiance-based daylight simulations for a test office with external blinds. Energ Buildings 33(7):683–697
70. Littlefair PJ (1990) Review paper: Innovative daylighting: Review of systems and evaluation methods. Light Res Technol 22(1):1–17
71. Andersen M (2002) Light distribution through advanced fenestration systems. Build Res Inf 30(4):264–281
72. SolaVeil. http://www.solaveil.co.uk. Accessed 2011
73. Lee ES, Selkowitz SE, Hughes GD, Clear RD, Ward G, Mardaljevic J, Lai J, Inanici MN, Inkarojrit V (2005) Daylighting the New York Times headquarters building. Lawrence Berkeley National Laboratory. Final report LBNL-57602
74. Selkowitz S, Lee ES (1998) Advanced fenestration systems for improved daylight performance. In: Daylighting '98 conference proceedings, Ottawa, ON, Canada
75. Lee ES, DiBartolomeo DL (2002) Application issues for large-area electrochromic windows in commercial buildings. Sol Energ Mat Sol C 71(4):465–491
76. Sullivan R, Rubin M, Selkowitz S (1996) Energy performance analysis of prototype electrochromic windows, Report number: LBL-39905 BS-360. Technical report, Lawrence Berkeley Laboratory
77. Inoue T, Ichinose M, Ichikawa N (2008) Thermotropic glass with active dimming control for solar shading and daylighting. Energ Buildings 40(3):385–393
78. Optics5. Lawrence Berkeley National Laboratory. http://windows.lbl.gov/materials/optics5. Accessed 2011
79. Apian-Bennewitz P, von der Hardt J (1998) Enhancing and calibrating a goniophotometer. Sol Energ Mat Sol C 54(1–4):309–322
80. Andersen M, Michel L, Roecker C, Scartezzini JL (2001) Experimental assessment of bi-directional transmission distribution functions using digital imaging techniques. Energ Buildings 33(5):417–431
81. Reinhart CF, Andersen M (2006) Development and validation of a Radiance model for a translucent panel. Energ Buildings 38(7):890–904
82. Andersen M, Rubin M, Powles R, Scartezzini JL (2005) Bi-directional transmission properties of venetian blinds: experimental assessment compared to ray-tracing calculations. Solar Energy 78(2):187–198
83. Greenup PJ, Edmonds IR, Compagnon R (2000) Radiance algorithm to simulate laser-cut panel light-redirecting elements. Light Res Technol 32(2):49–54
84. Mardaljevic J (2007) Climate-based daylight modelling for evaluation and education. In: VELUX daylight symposium, Bilbao, Spain, 6–7 May 2007
85. Breitenbach J, Lart S, Langle I, Rosenfeld JLJ (2001) Optical and thermal performance of glazing with integral venetian blinds. Energ Buildings 33(5):433–442

86. Carter DJ (2002) The measured and predicted performance of passive solar light pipe systems. Light Res Technol 34(1):39–51
87. Jenkins D, Muneer T, Kubie J (2005) A design tool for predicting the performances of light pipes. Energ Buildings 37(5):485–492
88. Reinhard E, Ward G, Pattanaik S, Debevec P (2005) High dynamic range imaging: acquisition, display, and image-based lighting (the Morgan Kaufmann series in computer graphics). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA
89. Wienold J, Christoffersen J (2006) Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energ Buildings 38(7):743–757
90. Painter B, Fan D, Mardaljevic J (2009) Evidence-based daylight research: development of a new visual comfort monitoring method. Lux Europa, Istanbul, pp 953–960
91. Newsham GR, Aries MBC, Mancini S, Faye G (2008) Individual control of electric lighting in a daylit space. Light Res Technol 40(1):25–41
92. Newsham GR, Arsenault C (2009) A camera as a sensor for lighting and shading control. Light Res Technol 41(2):143–163
93. Crisp VHC (1977) Preliminary study of automatic daylight control of artificial lighting. Light Res Technol 9(1):31–41
94. Hunt DRG (1977) Simple expressions for predicting energy savings from photo-electric control of lighting. Light Res Technol 9(2):93–102
95. Hunt DRG (1979) Improved daylight data for predicting energy savings from photoelectric controls. Light Res Technol 11(1):9–23
96. Crisp VHC (1978) The light switch in buildings. Light Res Technol 10(2):69–82
97. Bordass B, Cohen R, Standeven M, Leaman A (2001) Assessing building performance in use 3: energy performance of the probe buildings. Build Res Inf 29(2):114–128
98. Heschong Mahone Group (2006) Sidelighting photocontrols field study. Final Report to Southern California Edison Co, Pacific Gas & Electric Company and Northwest Energy Efficiency Alliance
99. DTI (2000) Photovoltaics in buildings: BIPV projects programme – strategy document 1998–2002. Technical report, Dept. of Trade and Industry
100. Yoon S, Tak S, Kim J, Jun Y, Kang K, Park J (2011) Application of transparent dye-sensitized solar cells to building integrated photovoltaic systems. Build Environ 46(10):1899–1904

Daylighting Controls, Performance and Global Impacts

Helmut Köster
Köster Lichtplanung, Frankfurt am Main, Germany

Article Outline

Glossary
Definition of the Subject
Introduction: What Is Daylight?
The Benefits of Daylighting: Energy, Health, and Design Flexibility
The Challenges of Daylighting
Designing Effective Daylighting
Conclusion
Bibliography

Glossary

Brightness contrast/glare Risks of glare in daylighting

- Direct glare: sun is in user's immediate field of vision
- Background glare: brightness contrast between monitor and monitor background
- Reflected glare: mirroring effect on monitor surface

Daylight autonomy Daylight autonomy of a workplace is the percentage of normal working time without the requirement of electric lighting – i.e., the time in which the target illuminance can be maintained by daylight alone. This varies depending on the minimum illuminance required and is determined using daylight coefficient.

Direct/diffuse light Diffuse light illuminates a room or area contrast or shadow reduced. It is usually caused by extensive light sources like the overcast sky (5,000–20,000 cd/m^2). In contrast the clear sky has a luminance of up to 50,000 cd/m^2 which is caused by the high illuminance of the direct sun (100,000 lx).

Efficiency of daylight Depending on the type of building, 20–40% of its total energy requirement is used merely for electric lighting, primarily during the day. Using optimized daylight redirection systems, electric lighting demands can be reduced to less than 10% of total energy loads.

Heat gain from daylight versus electric light Outside, daylight produces up to 120 lm/W of energy. Inside – behind glass with controlled solar heat gain coefficients (SHGC) – daylight offers even greater 240 lm/W. Fluorescent luminaires, on the other hand, can only achieve approximately 70 lm/W, requiring much higher energy use per unit of lighting provided. As a result, the heat gain in buildings using daylight is less than 1/3 of the

heat generated by electric lighting, a gain taken into account in the energy balance of a building through the SHGC value.

Illuminance (lux/lumen/candela) Basic terms of lighting technology are:

- Illuminance E [lx]. The total luminous flux incident on a surface per illuminated surface area
- Luminous flux Φ [lm]. The radiated power emitted by a light source or the radiated power incidence on a surface
- Luminous intensity I [cd]. The directional luminous flux of a light source

Light distribution curves (LDC) Light distribution curves of daylight systems indicate the direction and the intensity of the light distribution. Given the sun's changing angles of incidence outside, factors to be considered in mapping the light distribution inside include light redirection elements and the potential for adjusted tilt angles to ensure freedom from glare.

Quality assurance for daylight control Quality assurance must ensure the simultaneity of three criteria:

- Reduction of solar irradiation in summer (F_c value) for passive cooling balanced by solar gain in winter
- Sufficient daylight supply on the task surfaces (daylight coefficient)/visual comfort
- Sufficient visual transmission for quality views

Conventional louvers are generally closed to protect against glare and overheating. This not only hinders the view outside, but also prevents sufficient daylight from coming in. Design for daylight, including light redirecting louvers, ensures simultaneity of all three qualities.

Reflectors – contoured prismatic, mirror and diffusing blinds Basic functions of mirror or prismatic light redirecting louvers include reflection of the light back to its source to protect against overheating (thermal comfort) and/or redirection of daylight into the space to improve the illumination in the room (visual comfort). Precise contours make it possible to calculate exactly the quantitative light distribution (both inside and outside) with reference to the louvers' tilt angle and the altitude of the sun. Mono-reflective light control refers to the redirection of light inside and its deflection back outside in a single reflection in order to minimize its absorption by the blinds and any resulting heat development. Mono-reflective light control is required to prevent undirected reflection of the rays back and forth between louvers to prevent unplanned light scattering or absorption.

Shading for passive cooling balanced by passive solar heating Shading is a critical protective function against overheating to avoid active cooling, achieved exclusively through reflection of solar irradiation. At the same time, "passive solar heating" relies on solar transparency to achieve high levels of solar gain to reduce mechanical heating requirements. To avoid the risk of overheating in summer, it is critical to provide seasonally or daily dynamic devices that balance light and heat gain.

Total solar energy transmission (g/SHGC) Total solar energy transmission through the glazed area (g-value, SHGC) defines the heat load resulting from solar irradiation, light transmission, and secondary heat radiation ($g = \tau_{verg} + q_i$). Design decisions include the total energy transmission for the window assembly and its interior and exterior layers that form an overall system including sun and light control. The F_c value for the system indicates the percentage by which the total energy transmission through the glazing is reduced through redirection of the light: $F_c = g_{tot}/g_{glazing}$.

Definition of the Subject

Worldwide, efforts are made to introduce energy saving regulations and laws requiring the use of regenerative energies. The focus so far has been on energy conversion systems to obtain power and heat. Yet, daylight is one of the most substantive renewable energies, and highly underutilized in buildings. It should be obvious that using the sun for natural illumination of indoor environments, without overheating, should be a primary goal for the building sector.

Improving the supply of natural daylight in workplaces rather than using electric lighting is a highly sustainable and economic resource that saves both energy and power. Over 30% of the total energy consumption of average buildings is used for electric lighting, and most of this during the daytime! Simply by redirecting daylight into the room, the use of electric

lighting can be reduced by at least 50%, and effective shading with daylight offers further cooling energy savings. Daylight management systems are thus the focal point of the latest strategies to save energy and reduce the carbon footprint of buildings.

This entry outlines the critical characteristics of daylighting design, incorporating reflection, redirection, and diffusion in the design of façades and their external and internal layers. The correct manipulation and exploitation of the sun and daylight is vital as an energy savings resource and must mature into a key element in the development of energy concepts. Daylight design must become central to a range of disciplines, including lighting technology, façade technology, energy technology and architecture, and ultimately calls for integrated design studies to provide training in planning and coordinating between the various disciplines.

Introduction: What Is Daylight?

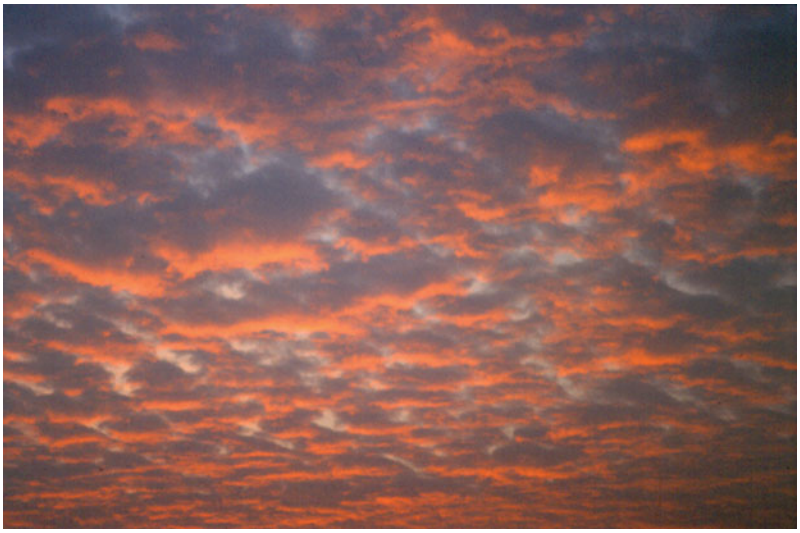
Common parlance and building physics define "daylight" as follows: Daylight is perceived as brightness outdoors and includes all variants from twilight to the brightest time of day when the sun is highest, ranging from diffuse light in shaded areas or with cloudy skies to direct sunlight. The sun is the primary source of all variants of daylight (Photo 1).

Solar radiation is approximately 1 kW/m^2 on the earth's surface. The illuminance from direct sunlight is approximately 100,000 lx. A clear, blue sky has approximately 50,000 cd/m^2, a cloudy sky may have less or more – depending on the time of year – between 5,000 and 15,000–20,000 cd/m^2 [17].

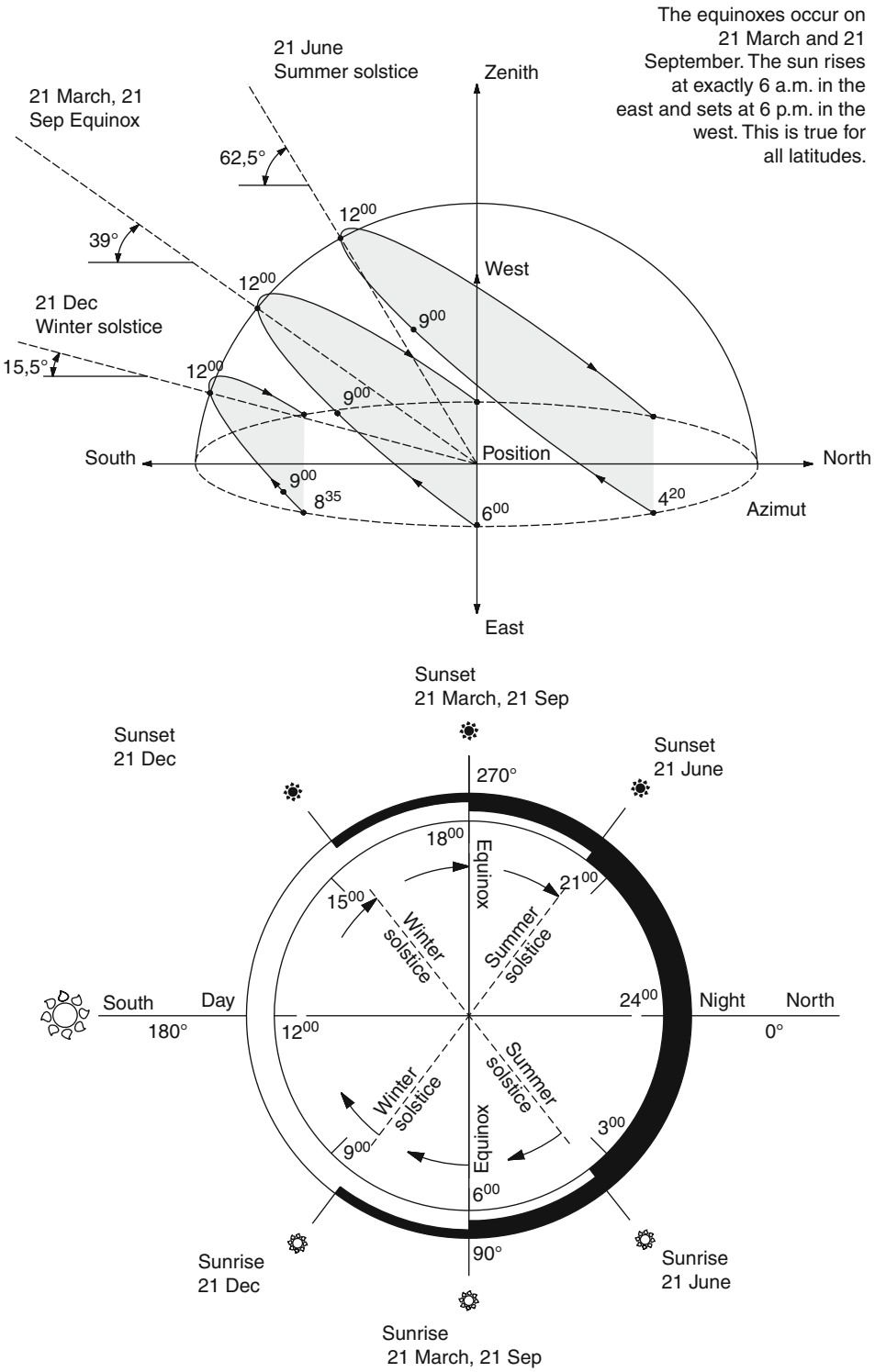
Depending on the location, daylight is in the wavelength range from 350 to 750 nm – roughly 40–50% of light radiation in the sensitivity range of our vision (Figs. 2, 3b–d, 6).

Whereas the term "daylight" more likely conjures up an image of diffuse light without directional reference – i.e., uniform light distribution from all directions – the sun itself has clear directional reference, which indicates to us the exact time of day and year (Fig. 1).

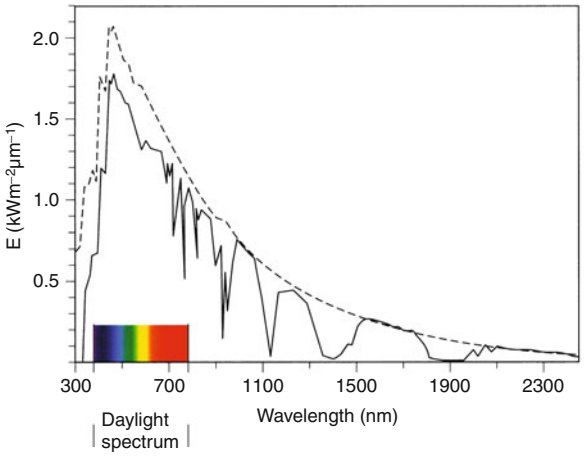
The number of daylight hours varies depending on latitude and time of year, and the number of sunny hours varies depending on the climate zone. Germany gets between 1,200 and 1,500 h of sunlight per year [5].



Daylighting Controls, Performance and Global Impacts. Photo 1
Daylight is an experience of color, season, and daytime and helps human beings to reintegrate themselves into the cycles of nature



Daylighting Controls, Performance and Global Impacts. Figure 1
Illustration of the daily and annual paths of the sun at 51° latitude [14]



Daylighting Controls, Performance and Global Impacts. Figure 2
Spectral energy distribution of daylight [9]

In many parts of California, the sun shines over 3,000 h [23] of approximately 4,300 daytime hours a year.

Design for daylight must address diffuse sky and sunny conditions, as they vary by orientation, season, and climate. Daylight design must address illuminance, glare, brightness contrast, and view content for visual quality in addition to managing solar heat gain for thermal quality.

Given the dynamic nature of daylight, light-transmitting building elements (windows and skylights) must be supported by "daylight technologies" to manage light transmission and redirection, balancing the variations in the sun's angles of incidence (Fig. 14) against variations in desirable heat energy transmission. The aim of daylight technologies is to achieve specific lighting effects using defined reflection, transmission, and/or absorption characteristics. Its further purpose is to gain more heat in winter and ensure shading for passive comfort in summer.

Effective daylighting design is dependent on the size of the windows relative to the proportion of the room, the configuration of the windows and their adjacent surfaces (to manage brightness contrast), the arrangement of windows relative to each other and the surfaces to be lit in the room (bi-directional and indirect light), the type of glazing (e.g., transparency, translucency, diffusion) including color effects (e.g., church windows) and the layers placed inside or outside of the glass that may change the transmission or directionality of the light.

Balancing Light and Heat: The Optimization of Daylighting

Current research and development in daylight-optimized glazing technologies concentrates on optimizing the energy transmission proportional to light transmission. Standard glazing, therefore, is defined on the basis of three physical reference parameters:

- Light transmission $\tau_{L \text{ or } VT}$ (Fig. 3d)
- Total solar energy transmission or solar factor (g or SHGC) (Fig. 3a)
- Heat transmission coefficient U in W/m^2K (Fig. 3a)

Whereas τ_L and g are parameters related to the transmission of light and heat from the outside to the inside, the U-value is used to calculate the conductive heat losses from the inside to the outside. To maximize daylighting year round, while avoiding solar heat gain as cooling loads in summer, the τ_L and g-values (SHGC) will be critical (Fig. 3).

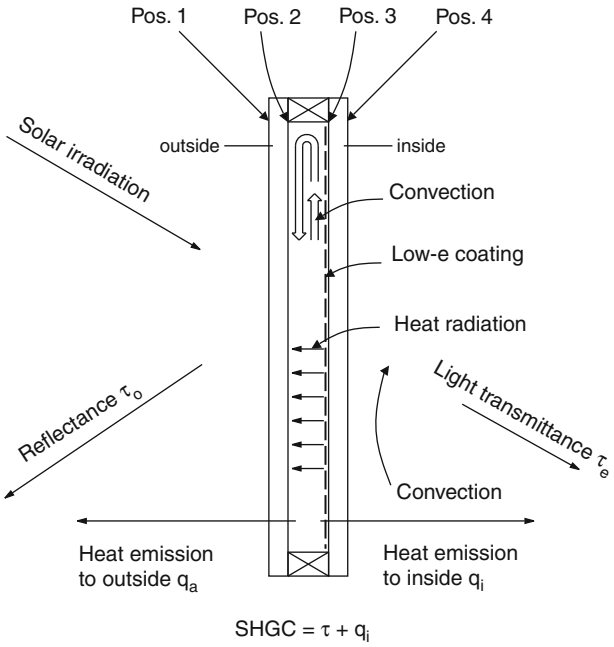
Thirty years ago, the visible light transmission (τ_L or $_{VT}$) of glass paralleled the heat transmission (SHGC or previously SC), with efforts to reduce heat transmission dominating specifications in an effort to minimize cooling loads. As a consequence, generations of dark or reflective glass buildings continue to be built, despite the fact that glazing innovations today allow for independent selection of visible transmission and shading values.

Typical glazing (6/16/6 and 4/12/4) of the latest developments has the following characteristics (Table 1).

The most recent developments in selective coatings applied to glass surfaces use sputtering methods to allow a ratio of τ_L/SHGC of approximately 2 for insulated (double-glazed) glass units. These advances ensure a high level of light transmission τ_L while keeping the total energy transmission g (SHGC) to a minimum. The competition among glazing producers can be reduced to the following formula:

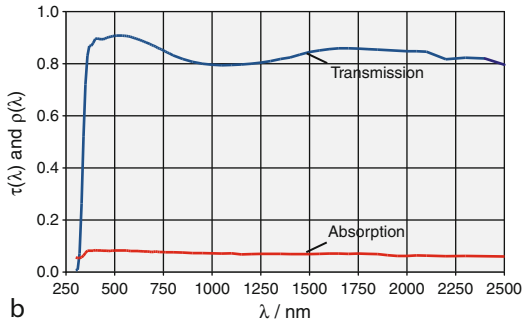
"More light, less heat"

In spirit, this formula is a vast improvement over less heat ergo less light – the reflective glass era.

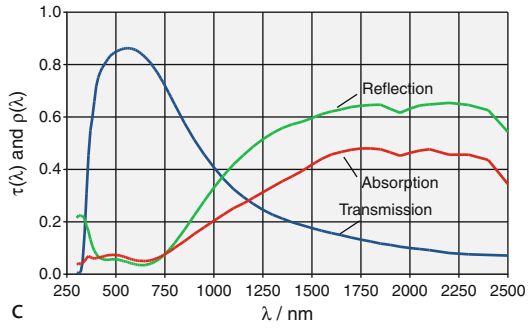


Typical design of insulation glazing in housing construction with low-e coating in pos. 3. (SHGC = 0.60)
The low-e coating of administrative buildings is found on pos. 2 (SHGC = 0.50)

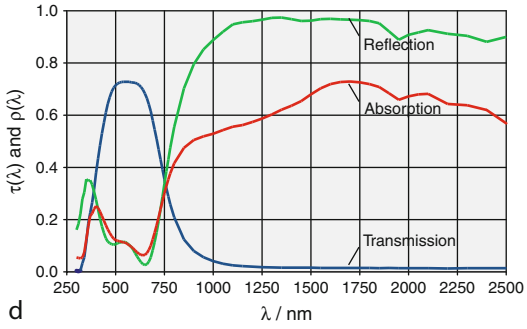
a U-value with double insulation glazing based on DIN 1.0 W/m^{2}K
U-value with triple insulation glazing based on DIN 0.6 W/m^{2}K



b Spectral coefficients of uncoated glass (outside)



c Spectral coefficients of low-e glass (inside)



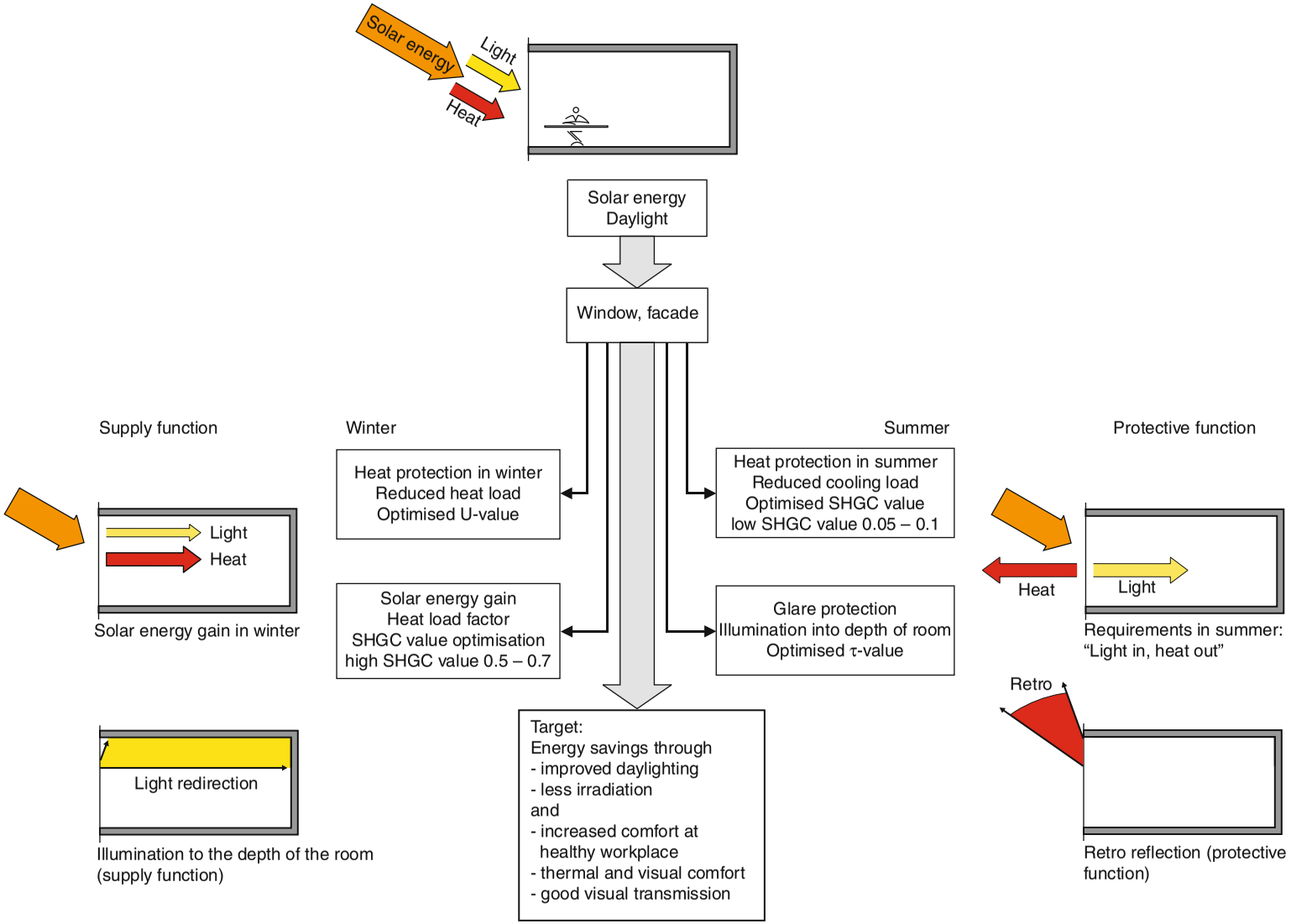
d Spectral coefficients of colour neutral sun protection glass (aussen)

Daylighting Controls, Performance and Global Impacts. Figure 3
(**a**–**d**) Basics of building physics in daylight technology [12]

Daylighting Controls, Performance and Global Impacts. Table 1 Light to heat ratio [19]

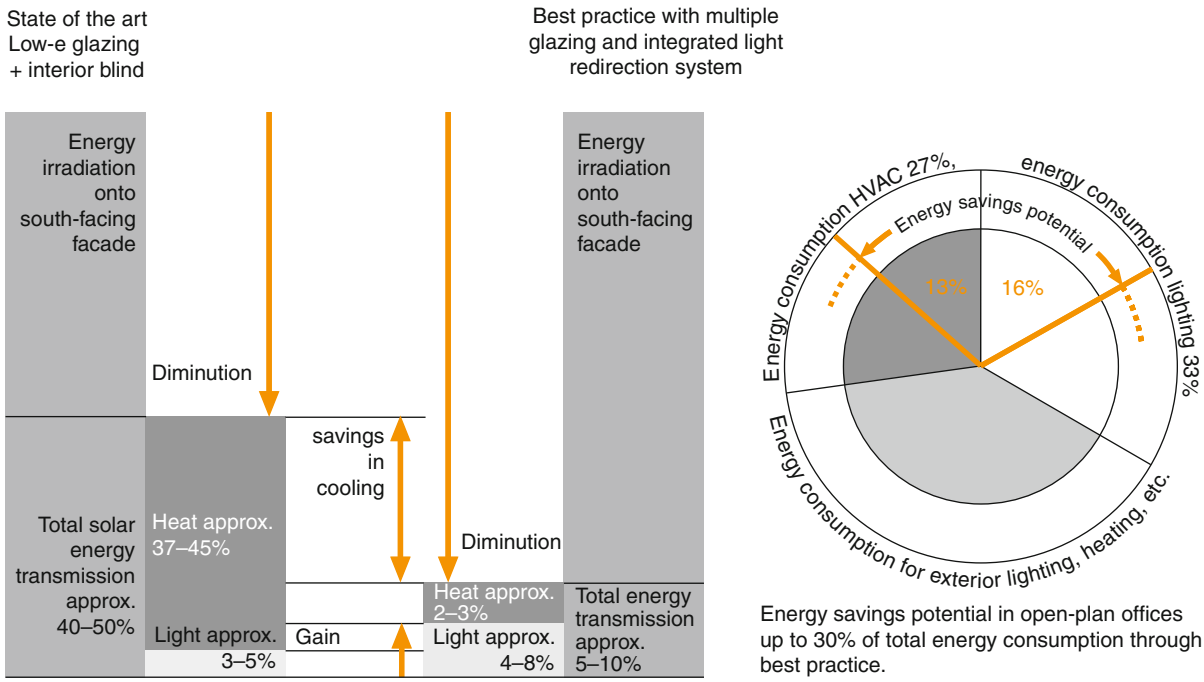
	τ_L in %	SHGC in %
Sun protection glass	53	28
	61	44
	60	28
Low-e glass iplus 3LS	80	63
	78	61
	71	42

However, there are at least two other climate conditions that demand a more nuanced selection in glazing or glazing controls. First, climates that have high heating loads in both residential and commercial buildings will not benefit from low solar heat gain coefficient glazing materials. The potential for passive solar heat as a natural or renewable heating source is very significant (Table 1). In colder regions, it is best to use glazing with a high light and high total solar energy transmission to realize the additional solar gains in winter. Second, climates that are hot and sunny, with significant

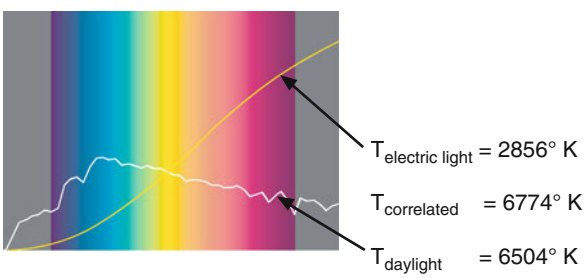


Daylighting Controls, Performance and Global Impacts. Figure 4

Tasks and functions of light-transmitting components in the conflict between required solar energy and passive cooling [14]



Daylighting Controls, Performance and Global Impacts. Figure 5
(**a**, **b**) Energy savings potential of a new, intelligent façade and daylight system in comparison to the state of the art

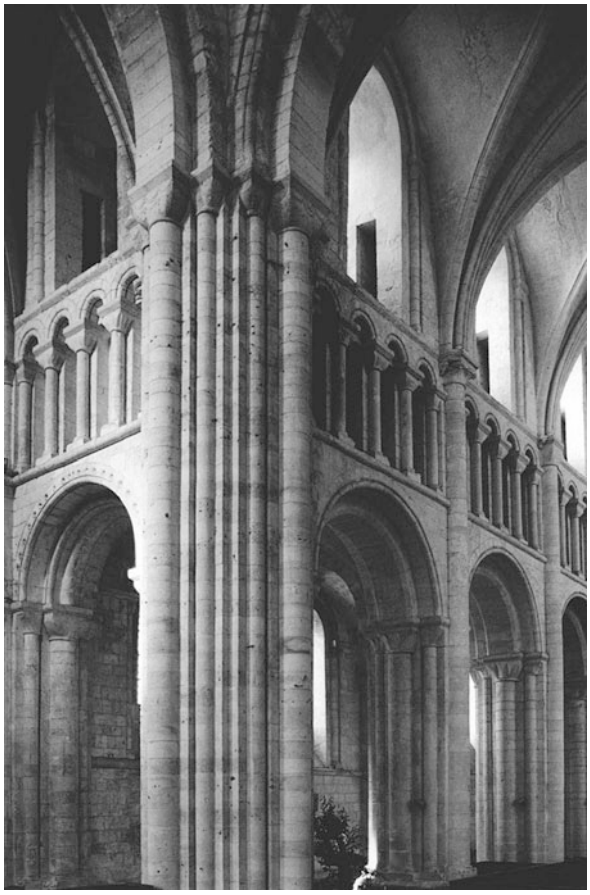


Daylighting Controls, Performance and Global Impacts. Figure 6
Spectral distribution of daylight [18]

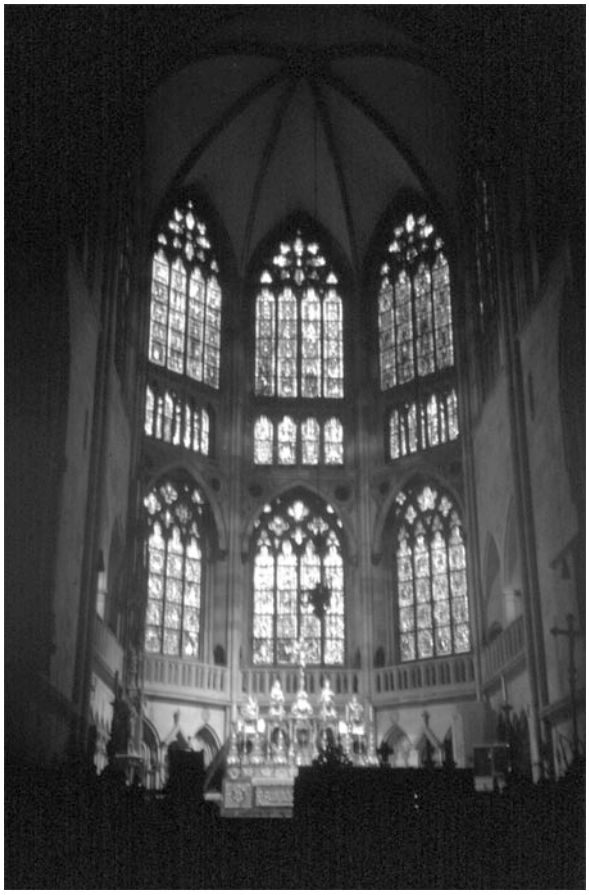
ground reflectivity, such as desert regions, have an additional challenge of managing both heat gain and brightness contrast, suggesting lower light and heat transmission glazing choices (Fig. 3d). Both of these climates will benefit from the lowest heat transmission coefficient possible (U-value) to reduce heat loss and heat gain (Fig. 4).

The highest level of natural conditioning through daylighting, shading, and passive solar heating can only be achieved through the addition of dynamic layers, either inside or outside the glazing, to allow for the exact amount of light and heat that is needed given time of day, season, orientation, and space function, in each climate.

The addition of dynamic layers becomes even more important as the design community pursues all glass buildings with floor to ceiling transparencies, regardless of the climate in which they are built. One only needs to see the latest buildings in Dubai, Hamburg, or New York to realize that façade design will need to resolve the transmission of daylight, solar heat and heat loss with dynamic layers, to enable natural energies to offset the major conditioning demands these façades have generated.



Daylighting Controls, Performance and Global Impacts. Photo 2
Examples from building history: Romanesque: interplay of wall relief and opening [14]



Daylighting Controls, Performance and Global Impacts. Photo 3
Examples from building history: Gothic: colored window [14]

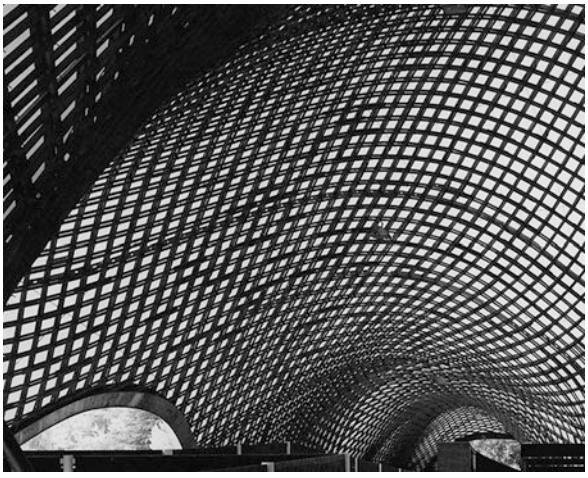
The Benefits of Daylighting: Energy, Health, and Design Flexibility

Energy Saving Potential of Daylighting

Electric lighting energy consumption [kWh] in conventional office buildings today is as much as 35% of the total electric load – demands that are generated primarily during the day when daylight is abundant [8]! Since the energy drawn for electric lighting is ultimately converted into heat, there is additionally a load on the cooling system. Proportional to the total energy used, electric lighting can add as much as 16% to the cooling energy bill, such that the combined electricity costs for lighting and cooling are almost 50% of total electric demand (Fig. 5). While total energy consumption is made up of both electricity and fossil fuel energy uses, daylighting alone can reduce total energy use by as much as 25–30%, one of the most cost-effective investments for energy and carbon savings worldwide.

In the USA, where electric lighting is becoming pervasive during the daytime, the New Buildings Institute calculates that daylight harvesting systems can generate lighting energy savings of 35–60%. According to the US Department of Energy, daylight-response controls of skylights have demonstrated lighting energy savings in warehouses of 30–70%, without consideration of the additional cooling energy benefits.

Daylighting Controls, Performance and Global Impacts. Photo 4
Examples from building history: Baroque: indirect daylighting [14]



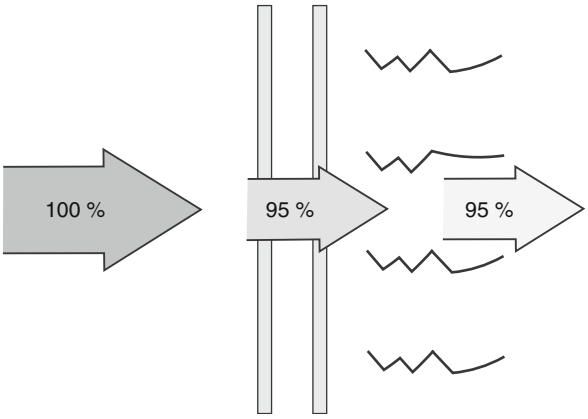
Daylighting Controls, Performance and Global Impacts. Photo 5
Examples of daylighting in modern architecture: Frei Otto: multipurpose hall and restaurant

Daylighting Controls, Performance and Global Impacts. Photo 6
Examples of daylighting in modern architecture: Jørn Utzon: church at Bagsvaerd

In a controlled experiment, the Energy Center of Wisconsin measured the additional cooling energy savings at 25% and fan energy savings at 3% (NBI, DOE, Wisconsin 2005).

The economic impact of ignoring daylight is even more problematic because it is an electric load in buildings – for which source or primary energy costs are significant. One kilowatt of power on site uses approximately 3–4 kW of primary energy, with the rest lost as heat up the chimney at the power plant. In conventional coal or oil fired power plants, only 35–40% of the primary energy is converted into power with a further 6% of the energy produced at the power plant lost in transmission [21]. In developed economies such as the USA, Japan, and Germany, power plants are to be blamed for approximately 50% of all CO_2 emissions! Over 40% of each nation's total energy consumption in developed economies is used for heating, cooling, air conditioning, lighting, and other power requirements in buildings [11].

Daylighting Controls, Performance and Global Impacts. Photo 7
Examples of daylighting in modern architecture: Paul Rudolph: Interdenominational Chapel



Daylighting Controls, Performance and Global Impacts. Figure 7
Color rendering of low-e glass 95–96%, color rendering of Retro louvers >99% eliminate?

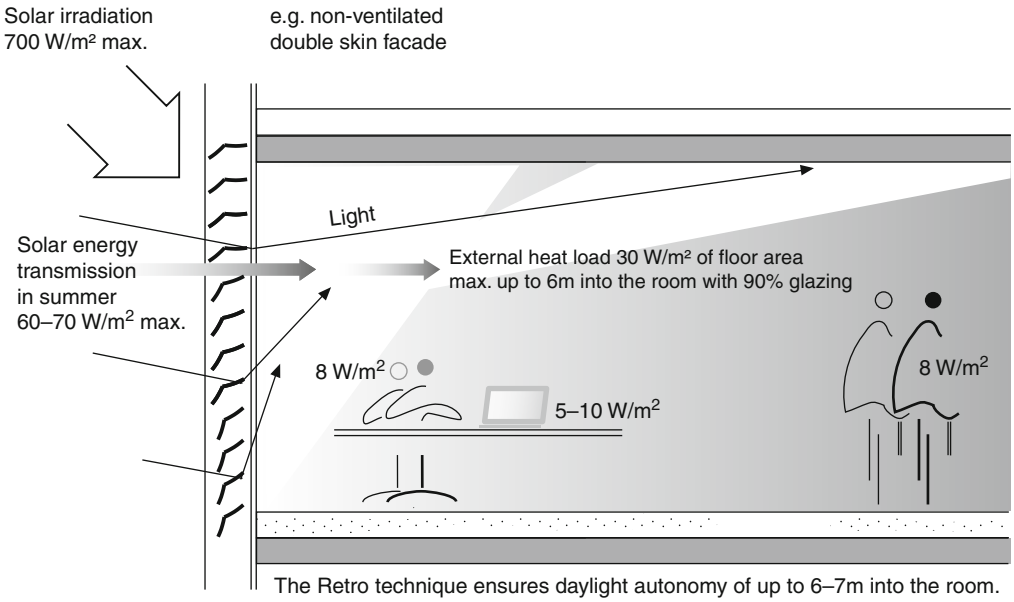
Daylighting Controls, Performance and Global Impacts. Photo 8
Color shift through sun protection glass and electrochrome glazing

Daylight and Health

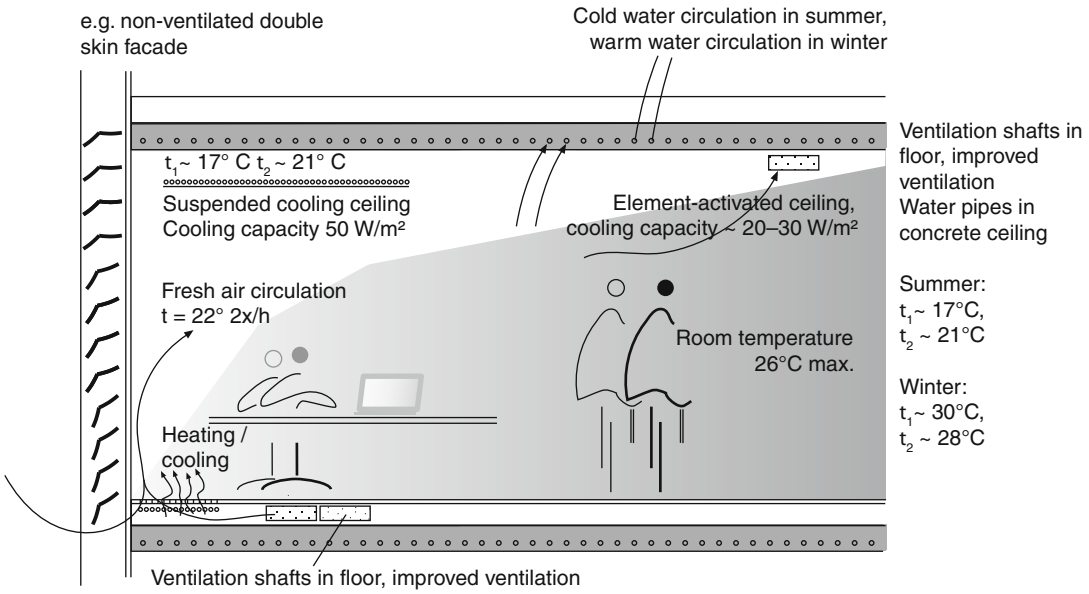
Humans, like plants, need to live and work in the full spectrum of light provided by [4] (Fig. 6). Daylight is critical for natural vitamin and hormone production, and is associated with numerous emotional values as well [9]. While the research on "daylight nutrient ingestion" through the human eye and the skin is still sparse, it is known that lethargy in winter and certain eating and hormonal disorders are due to a physiological lack of daylight [9]. These deficiency syndromes can be countered by using light treatments during which the patient stays, for hours, in "full-spectrum" settings for health.

One quality of full spectrum daylighting that is significant is its color rendering index. For this reason, it is important not to distort the natural color–rendering qualities of daylight by tinted glazing or colored blinds, screens, and fabrics that change the color composition of natural daylight (Photo 8, Fig. 7). As a result, quality assurance must also take into account the color-rendering index of the daylight transmitted through the window and blind assembly (Fig. 7).

These findings call for "healthy" building design guidelines that ensure full "value-added" daylight to the occupants of the building, changing the color composition of the daylight as little as possible. It is not enough to simply provide sufficient lighting for a visual



a External and internal heat loads in summer



b Heating and AC system concept

Daylighting Controls, Performance and Global Impacts. Figure 8

The Retro technique dramatically reduces the high outer heat loads, making possible new concepts in AC systems. As a result, chilled ceilings are sufficient in most climates to cool a building. (**a**) External and internal heat loads in summer. (**b**) Heating and AC system concept

Daylighting Controls, Performance and Global Impacts. Photo 9

Mirror façades darken the interior, making it necessary to switch on the lights during the day (Photo: Helmut Köster) [21]

task; daylight designers must ensure that each individual receives the necessary light nutrition in their workplaces.

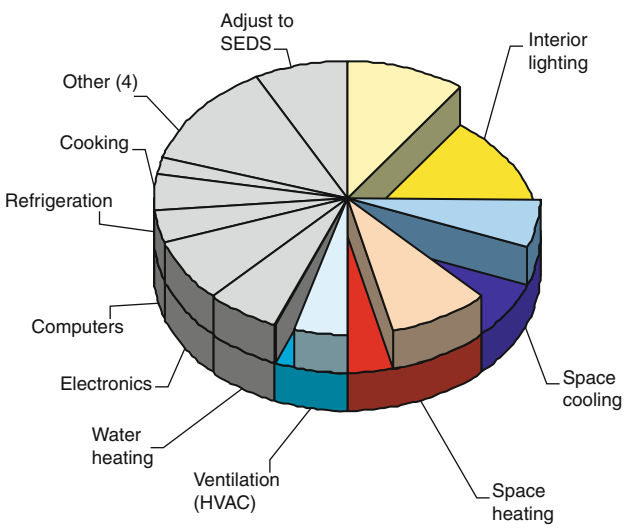
Daylighting and Design Flexibility

Effective daylighting design that manages cooling demands can also save building costs. The need for ceiling plenums are often driven by the demand for air conditioning, with air-based systems dominating the USA and now Asian building growth. The elimination of deep ceiling plenums for air-based cooling can save floor-to-floor heights, overall building height, and even associated elevator demands (Fig. 8).

Since managing solar heat and light can dramatically reduce cooling loads, the potential of radiant water-based cooling systems is unleashed. The use of water-based chilled ceilings, chilled beams, or radiant ceilings and walls for thermal cooling offers significant energy savings over air-based cooling [16]. In addition, the ducted systems can be dedicated to the delivery of fresh air, at 10% of the volume of air-based cooling systems. The ventilation-only systems can be focused on the quality and quantity of fresh air needed in each space, instead of ventilation that is dependent on the variability of cooling demands. Ventilation effectiveness is a measure of the delivery of outside air to the

	Consumption	Consumption
	Standard	Best practice
Interior lighting	25.2%	10.1%
Space cooling	12.5%	5.5%
Space heating	12.2%	8.5%
Ventilation (HVAC)	6.2%	4.7%
Water heating	6.0%	6.0%
Electronics	7.6%	7.6%
Computers	3.9%	3.9%
Refrigeration	4.2%	4.2%
Cooking	1.9%	1.9%
Other (4)	12.6%	12.6%
Adjust to SEDS	7.7%	7.7%

Savings in energy consumption: 27.3 %



Daylighting Controls, Performance and Global Impacts. Figure 9

Energy consumption savings of approximately 27.3% in the climate of Sofia in office buildings by the use of intelligent daylight systems which help to reduce the cooling loads in summer and simultaneously improve the daylight transmission

Daylighting Controls, Performance and Global Impacts. Photo 10
Prisms are translucent but prevent visual transmission [14]

occupant, and many cooling-dominated ventilation systems have compromised ventilation effectiveness. The separation of thermal conditioning and ventilation offers both energy benefits and air quality benefits.

The savings in building and operating costs far exceed the investment costs for new daylight technology and the investments in façades with improved heat protection.

The Challenges of Daylighting

Despite generations of successfully daylit buildings, designing for effective daylighting poses a number of challenges. First, daylighting must be balanced against overheating due to solar gain, a time of day and seasonal design challenge. Second, daylighting is only successful if the brightness contrast of the window or other daylight source is fully managed, and direct glare must also be avoided. Heat loss and heat gain must be managed through the window and skylight areas, since they are typically less resistive to heat transfer than today's wall and roof constructions. Finally, view must be considered a critical component of effective daylight design.

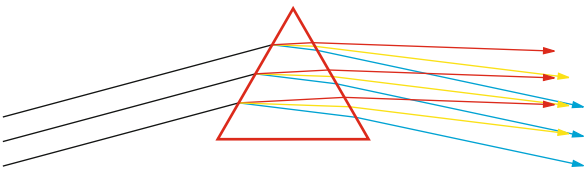
Daylighting with Effective Shading

Over the past 40 years, the predominant strategy for shading to reduce solar energy transmission into the

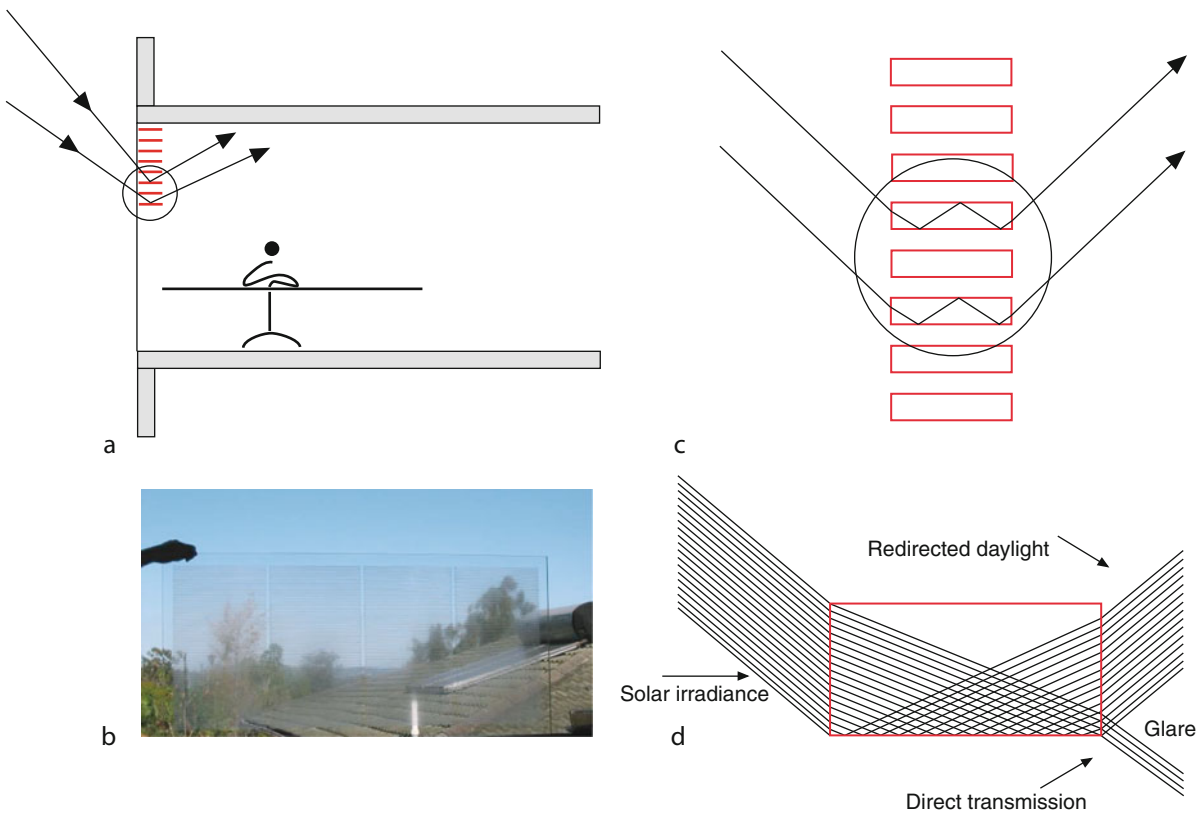
Daylighting Controls, Performance and Global Impacts. Photo 11
Prisms are translucent but prevent visual transmission [14]

building is the selection of highly reflective or mirror glass. The low solar heat gain glass reduces the solar load by reflecting the solar gain at the outer skin with a measurable level of absorption in the glass as well. These low solar heat gain coefficients traditionally ensured low daylighting transmission as well, failing to provide the buildings with sufficient natural daylight. The resulting increase in energy use for electric lighting and associated cooling ensured that the glazed area in buildings had a negative energy balance (Fig. 29).

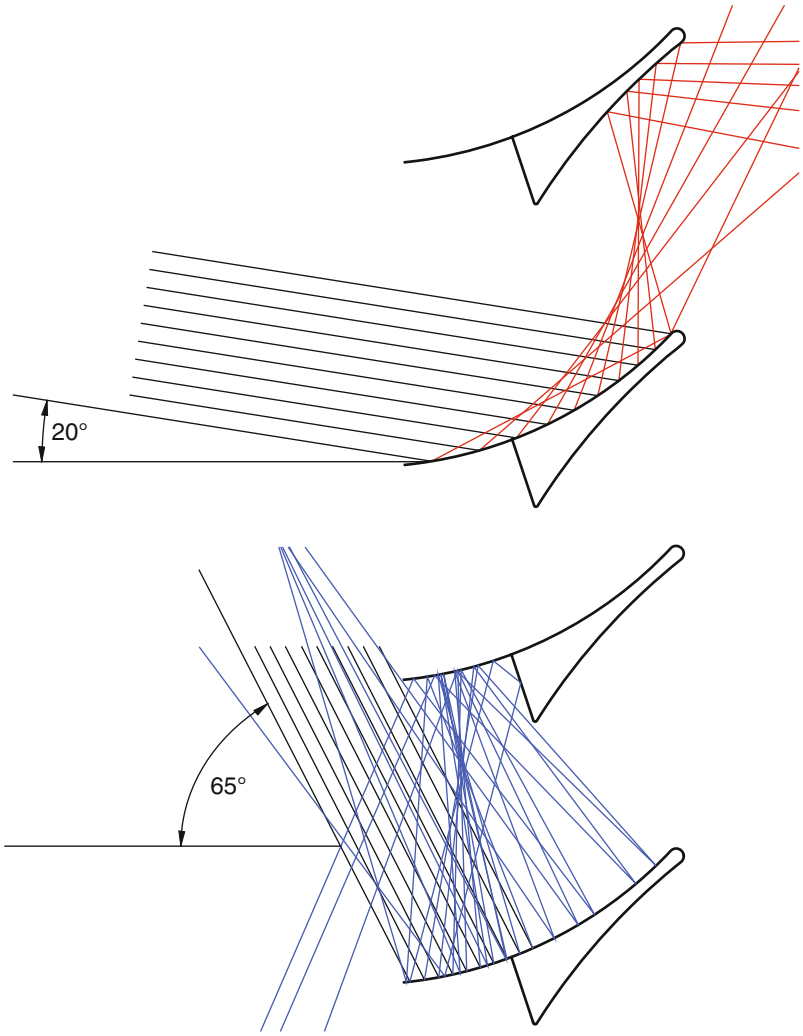
These sun protection strategies, which shade the inside but at the same time darken it to the point of requiring additional electric lighting even while the sun is shining, are, a priori, energetically highly counterproductive (Photo 9). While the use of reflective glass is the most problematic, many external and internal sun protection devices follow the same rule – darken the space to the point where electric light is necessary. This includes, in general, all interior roller shading systems, vertical louvers and colored blinds, and even many external sun shading devices. The loss of daylighting becomes even more profound when internal or external shading layers are used in combination with low transmission glazing.



Daylighting Controls, Performance and Global Impacts. Figure 10
Breakup of light into its spectral colors through refraction in the prism



Daylighting Controls, Performance and Global Impacts. Figure 11
Light deflection using laser-cut panels, advantage: good visual transmission, disadvantage: risk of overheating

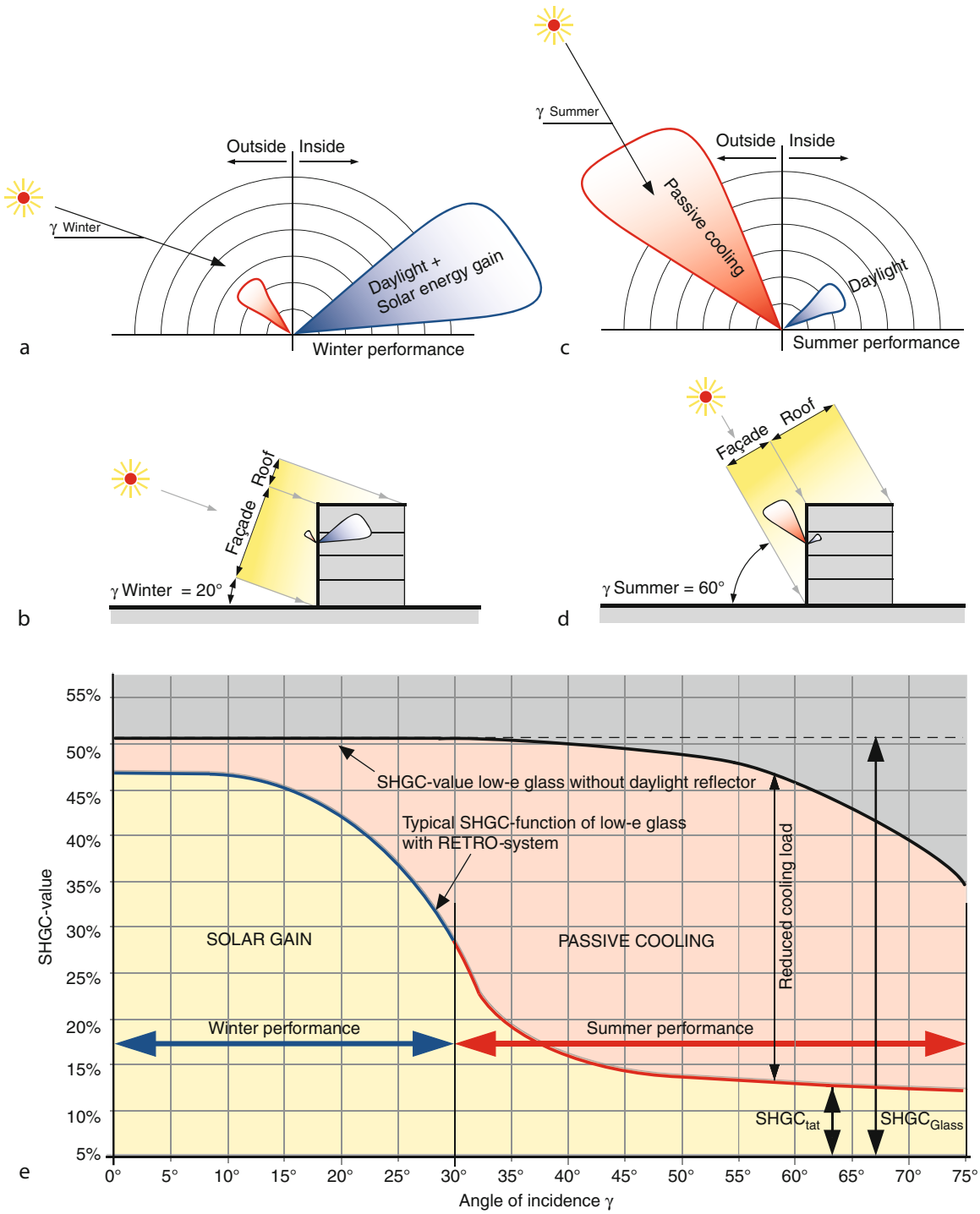


Daylighting Controls, Performance and Global Impacts. Figure 12
Okasolar light reflection louvers in a fixed position inside the glazing with resulting light redirection of the low winter sun and the high summer sun. A next step in the development of the systems involves mono-reflective light deflection which drastically improves the shading capacity of the technology (see also Fig. 16) [EP 0029442; US 4,715,358]

Shading strategies that do not include specific light redirection or light transmission solutions to improve the natural illumination in the room are a waste of the natural, free resource of daylight and of valuable electricity. Shading measures inside the glazing have further concerns relative to protecting the window zone from overheating, since they convert a percentage of the incident sun into heat through absorption. This is especially pronounced when the internal shades or blinds are dark, generating measurable cooling loads.

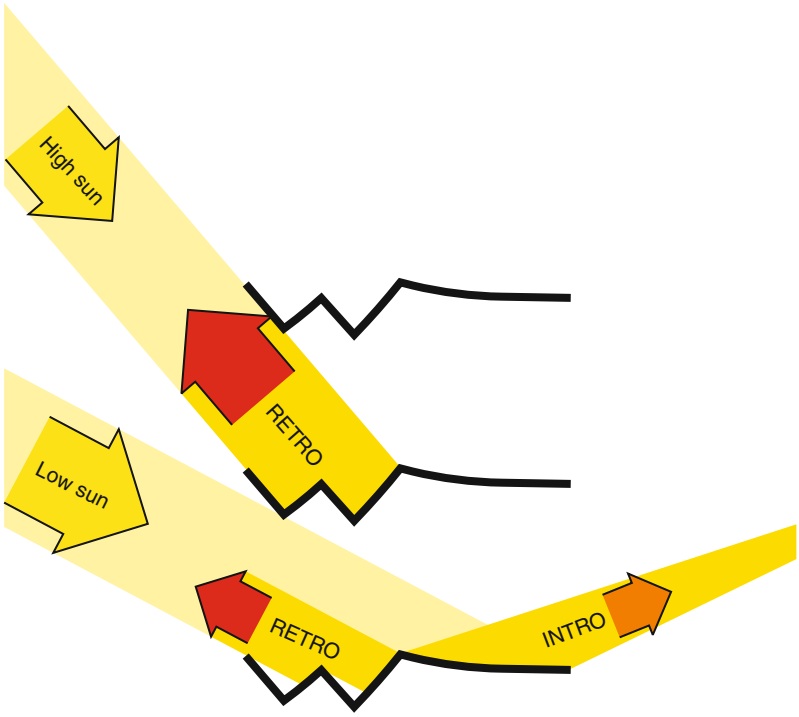
The pervasive loss of daylight as the dominant light source in modern buildings establishes the challenge to develop new strategies in ecological and sustainable building construction. A focus must be given to improving and optimizing the use of daylight without overheating the buildings through solar irradiation.

The first step is the introduction of high visible transmission glass combined with low solar transmission for any building that does not need passive solar heat. In cooling-dominated climates, this can be through high-visible, low-solar glazing materials.



Daylighting Controls, Performance and Global Impacts. Figure 13

(**a**–**d**) Thermal loads for roofs and façades. (**e**) Given the distinctive heating and cooling periods in the European climate, intelligent daylight technologies can maximize solar energy gain in winter and achieve effective shading for natural cooling in summer. These two conflicting goals are achieved through utilization of the sun's changing angle of incidence even without tilting of the blinds



Daylighting Controls, Performance and Global Impacts. Figure 14
Light redirected on a RetroLux louver (see also Fig. 23). The overheating summer sun is retroreflected on the horizontal, open louvers. Daylight from diffuse sky and a low sun is supplied via the light shelf. [US 6,240,999; US 6,845,805; EP 0793761; EP 1212508]

However, in mixed or heating dominated climates, the provision of daylight must be accompanied by dynamic shading to ensure the ideal seasonal or even time-of-day energy management. The state of the art for buildings in Europe is low-e glass with a light transmission of 80% and a solar heat gain coefficient or g-value of 55% for year-round daylighting and winter solar heating. This is then combined with dynamic external and internal sun protection for summer SHGC-values as low as 8–15%. Technically, it is possible to lower the SHGC values for the high, overheating sun to less than 5% while still maintaining effective daylighting. These energetic advantages are primarily achieved through light redirection using reflective surfaces within a special, double skin façade (Figs. 5, 19, 31, 32).

It is critical to remember that a building will heat up less in summer as a result of improved daylighting when compared to the use of electric lighting combined with low transmission glazing or glazing assemblies. The sun produces a photometric radiation equivalent of approximately 100–120 lm/W. With low solar/high visible transmission glass assemblies, the lighting energy increases to 200–240 lm/W. On the other hand, the lighting efficacy of fluorescent luminaires is only 60–70 lm/W. Given comparable illuminance solutions (lux [lx = lm/m^2]), buildings are heated three times as much by fluorescent lights as by daylight [2].

Effective Daylight Distribution with Brightness Contrast Control

Admitting daylight through high-visible transmission glass is only the first challenge. The distribution of that light to task surfaces, as well as the balance of brightness of various surfaces to ensure managed brightness contrast, may be the greater challenge (Figs. 21, 23, 24, Photos 18, 28).

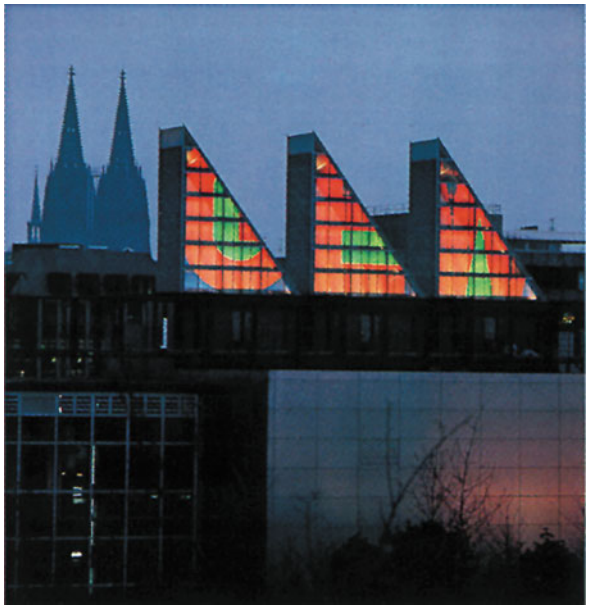
There are numerous textbooks on the design of window walls in relationship to room proportions

Daylighting Controls, Performance and Global Impacts. Photo 12
Holograms for light deflection produce excellent design effects using artificial lighting [14]

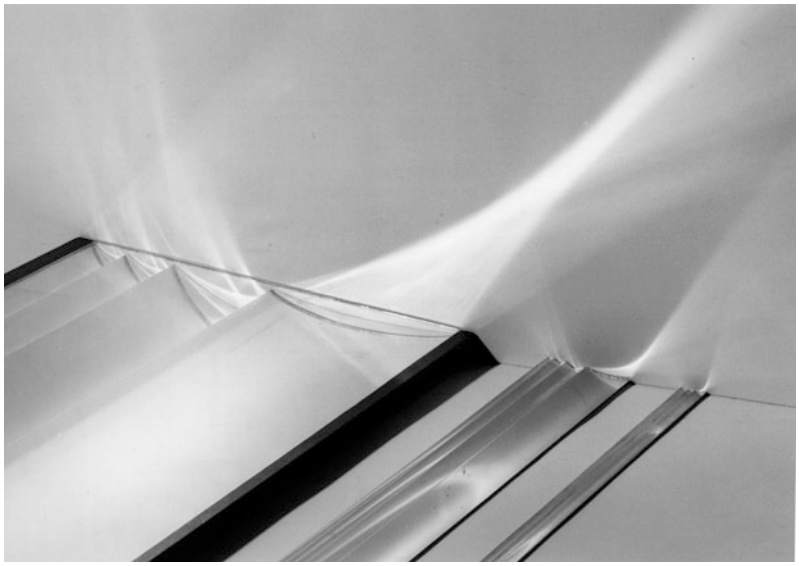
and surfaces for effective daylight distribution and brightness contrast control [20]. This entry does not intend to replicate these guidelines, or explain critical diagrams that illustrate the breadth and depth of these design imperatives. It is critical to state, however, that the design solution set will be significantly driven by climate and building function, and will have changing demands based on season and time of day.

Existing daylight design guidelines do not typically resolve the need for greater depth in daylight penetration for today's deeper buildings – to be achieved without glare or overheating. For this reason, it is critical to integrate dynamic controls of the daylight source – the windows and skylights – to ensure effective daylighting and control for solar heat gain. Dynamic

Daylighting Controls, Performance and Global Impacts. Photo 13
Holograms for light deflection produce excellent design effects using artificial lighting [14]



Daylighting Controls, Performance and Global Impacts. Photo 14
Holograms for light deflection produce excellent design effects using artificial lighting [14]



Daylighting Controls, Performance and Global Impacts. Photo 15
Light deflection using RetroLux™ louvers (see Fig. 7). Optical systems can be reduced or increased in size. The picture shows a 50-mm-wide louver, a 200-mm louver, and a large 2,500-mm louver. It illustrates how the high summer sun is reflected back outside while daylight is redirected to improve the illumination inside

Daylighting Controls, Performance and Global Impacts. Photo 16
Head office of Energie AG, Linz, Austria, Architects: Weber Hofer Partner, Zurich

controls for light distribution and shading are typically external or internal layers on the glazing, though they can be integral to the glass assembly. In designing the materials and surfaces of these layers, the following light redirection methods are important for consideration: prisms, holograms, mirrors, and hybrid systems.

Prisms redirect light through refraction in an optically denser medium, e.g., in acrylic or polycarbonate. Through total internal reflection, it is possible to prevent the light from passing through altogether due to retroreflective properties. The disadvantage of prisms is their lack of transparency. The systems are only transparent to light, offering translucency but not clear views (Photos 10, 11). When daylight passes through prisms and holographs, it is broken up into its spectral colors often creating colored patterns and striations on the walls and ceilings (Fig. 10). On the other hand, the use of laser-cut prismatic panels strategically placed on the inside or outside could ensure excellent views in combination with daylight redistribution (Fig. 11b). This installation, however, does not protect from overheating (Fig. 11c, d) [22].

Daylighting Controls, Performance and Global Impacts. Photo 17
Head office of Energie AG, Linz, Austria, Architects: Weber Hofer Partner, Zurich

Holograms are also useful for redirecting light into the depth of the room (Photos 12, 13, 14), but also eliminate transparency for views. They can be combined with electric lighting to offer some dramatic design effects in addition to effective daylight distribution. The foils onto which holograms are imprinted and which are embedded in glass, however, are still expensive, which makes it difficult to use them in larger quantities [13].

Mirror systems provide for a wide range of light redirection effects, depending on their mirror geometry, and they can be used on louvers or blinds inside the glass, within double skin façades (Figs. 31, 32, Photos 16, 17), outside the glass, or even installed in miniaturized form in the gap between the glass layers (Fig. 24, Photos 26, 29, 30).

Designing Effective Daylighting

The design of effective window and skylight systems for light and heat management as well as views requires consideration of a number of design variables: façade orientation and louver geometry, louver surface qualities and their relationship to ceiling diffusers, louver controls, and electric lighting interfaces.

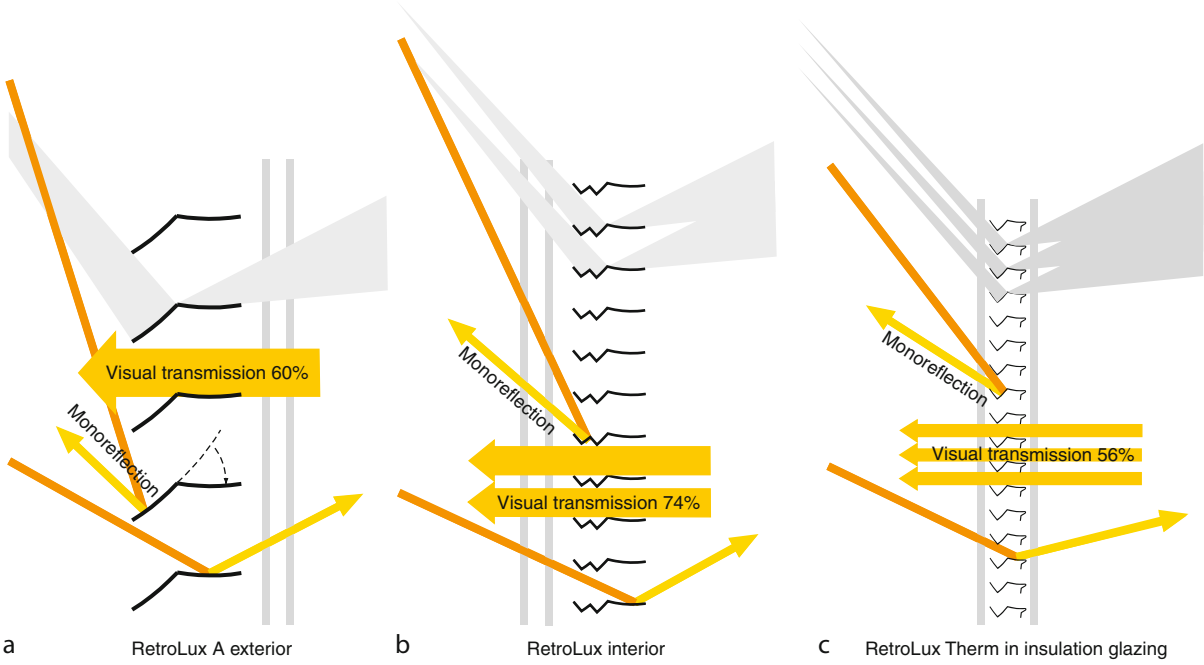
Designing Daylighting: Orientation and Basic Louver Geometry

Internal, integral, or external louvers and blinds offer the most strategic solution to meeting the dynamic demands for balancing daylight distribution with effective shading. The surfaces of these louvers are critical to daylight distribution, but so is the louver geometry.

Effective daylighting requires varying design responses for different orientations. Japanese design standards for commercial buildings are focused on only providing occupied spaces with south and north orientations to ensure effective daylighting without the overheating or glare that is prevalent on the east and west (Fig. 1). Roofs are subject to much higher energy loads in the summer than east west façades, and significantly more than southern and northern façades. In short, design for effective daylighting without glare and overheating will require unique detailing for each orientation.

The first step to optimizing louver geometry is to recognize that blinds should be inverted from their conventional downward facing arc to an upward facing arc intended to move daylight further into the space, working in combination with the ceiling as a reflector (discussed in a later section).

The second step to optimizing louver geometry is to include the elevation angles of the sun in the design of even fixed systems, appropriately controlling thermal energy and light transmission by season. The sun in winter shines at a lower angle and in summer the angle is significantly higher. Consequently, it is necessary to design a louver geometry which redirects the low winter sun to a greater degree into the building, while reflecting the high summer sun to a greater degree back outside. This is in order to achieve a homeostatic



Daylighting Controls, Performance and Global Impacts. Figure 15
RetroLux systems ensure simultaneity of retroreflecting the overheating summer sun on the retroreflector, redirecting diffuse daylight on the light shelf and good visual transmission

balance of the building in accordance with the heating and cooling periods in different seasons (Fig. 15).

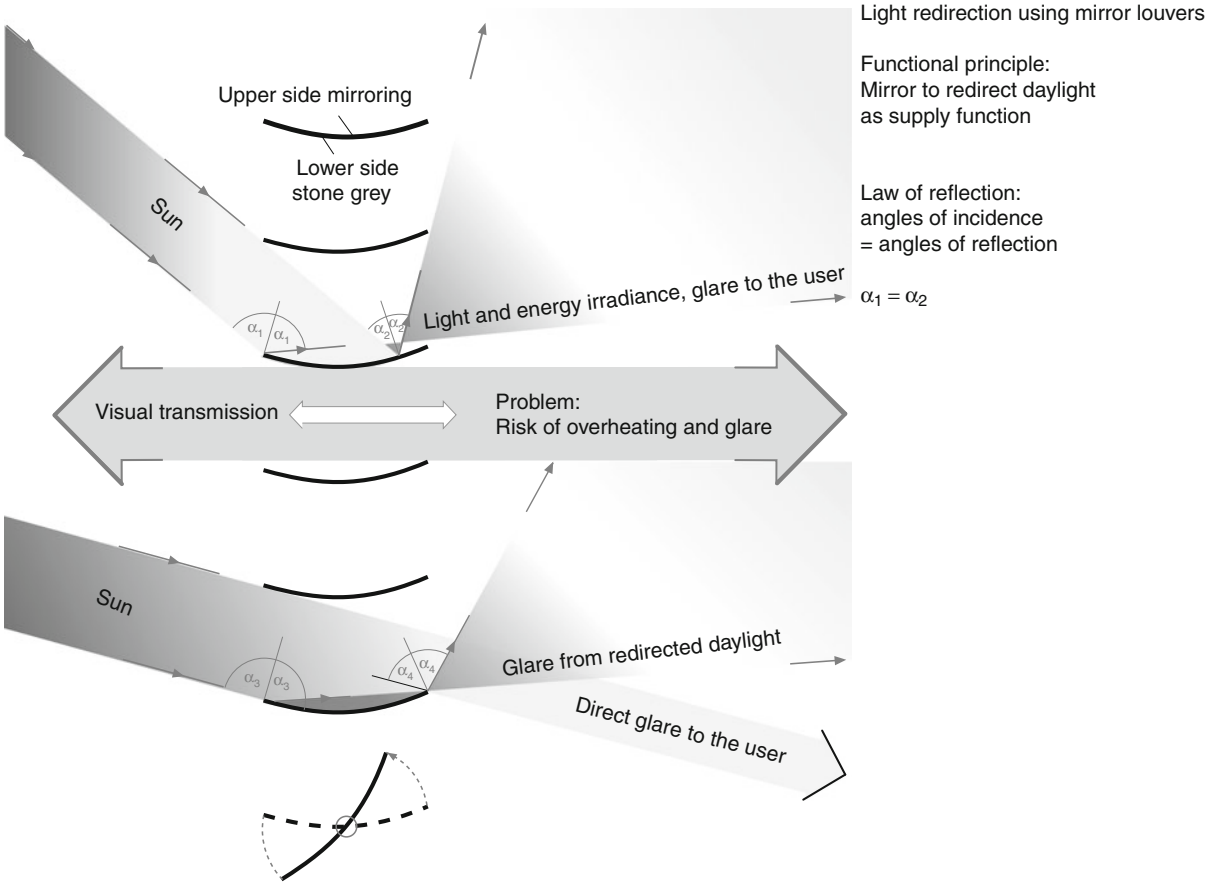
The development of light redirection systems is dependent on defining the friend/foe relationship with the sun in different climates and cultures: How much of the solar irradiation should be reflected back to the sky for passive cooling comfort? How much should be directed inside for daylighting of the interior? How much can the amount directed inside increase if passive solar heating is desired (Table 1, Fig. 13)

Such complex, contradictory demands for controlling energy transmission require precise mirror geometries. Horizontal louvers are most useful for refining energy transmission because they can respond to the solar altitude angle even without adjusting the louver position, and because they are best suited to ensure illumination into the depth of the room in combination with the unobstructed ceiling plane (Fig. 24). Vertical louvers or blinds, on the other hand, only respond to the azimuth of the sun without redirecting light toward the ceiling. They are merely suited to provide shade. The benefit of even fixed light redirection louvers with geometries optimized for summer shading is tremendous, as shown in Figs. 13 and 14.

Designing Daylighting: Louver Surface Qualities

Blinds should be designed both as shading devices and as light shelves. Maximizing the use of daylight with well-managed and comfortable illuminance levels, while minimizing solar overheating, can best be achieved through the louver contour and their optical characteristics. While mirrored surfaces are very effective for reflecting daylight deeper into the space, the surfaces of blinds can be significantly more refined to differentiate between high sun angle summer sun and low sun angle winter sun. Precise, mono-reflective surfaces can redirect daylight toward the ceiling and into the depth of the room.

Precise guidance of the daylight using optimized mirror geometries can also prevent the bottom sides of the louvers from exposure to sunlight, which would produce glare and turn the blind into an unwanted



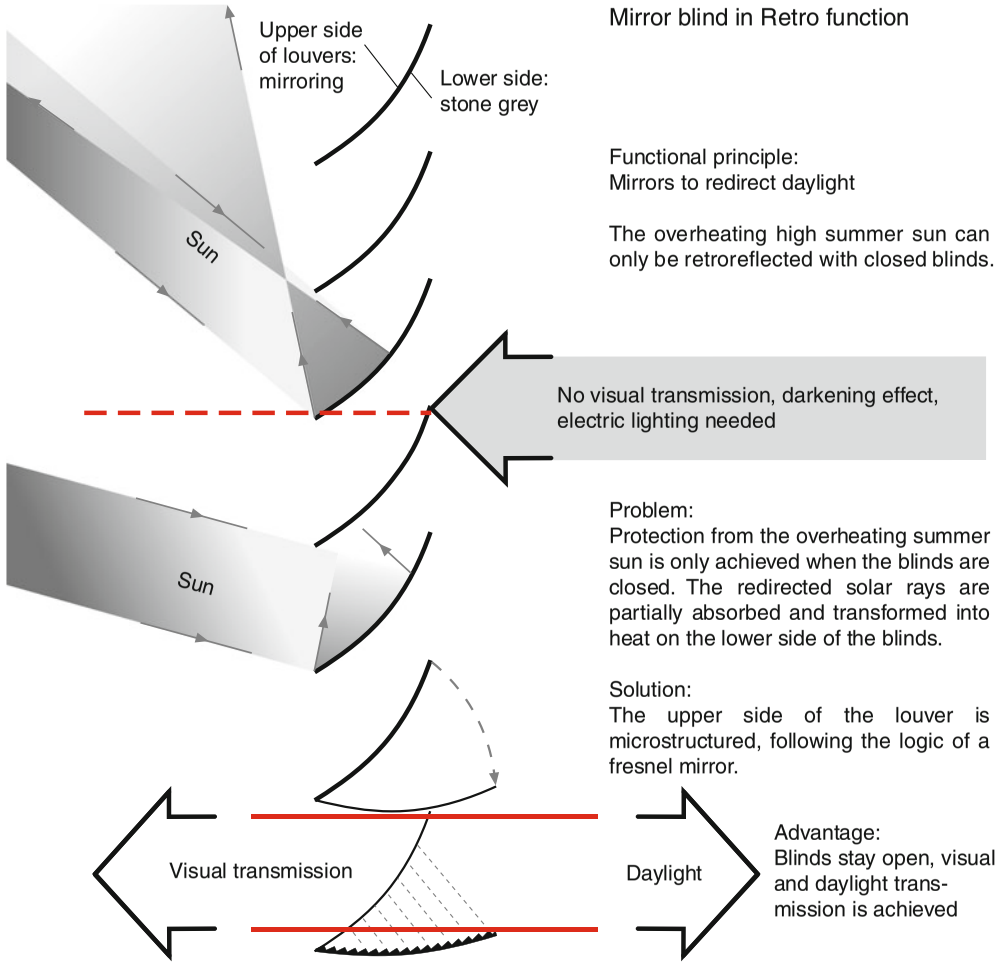
Daylighting Controls, Performance and Global Impacts. Figure 16
Mirror blinds open

surface emitter or even a heat radiator. Only mono-reflective structures ensure that incident sunlight is redirected either into the depth of the room and/or back outside, without a ping-pong effect between the louvers themselves (e.g., Fig. 12). Mono-reflective systems are easier to optimize in terms of their energetic and lighting capacity, and their thermal performance can be precisely calculated even in the building simulation phase (Figs. 18, 20, 23, 24). In addition, the underside of the blinds should not be mirrored, to avoid reflective glare.

Designing Daylighting: Innovative Louver Geometries and Surfaces

There are dozens of macro- and micro-design choices that can make a horizontal blind into an even more effective light fixture – with both glare and heat control. Surface reflectivity/absorptivity and finish, geometries of the outer and inner surfaces, W- and V-shaped blinds in addition to curved blinds with micro-features such as Fresnel surfaces, and innovative assemblies will be discussed further (Fig. 15).

If conventional blinds are inverted and mirrored on their concave upper side, they will indeed redistribute daylight to the ceiling and as a result to deeper portions of the room. However, the lower blinds will first direct light to the occupant, creating glare that makes it necessary to close at least the lower section of the blind (Fig. 16). As a result, visual contact with the outside is lost, and if the upper portion of the blind cannot be controlled independently, effective daylight will be lost. Even in the upper section of the window, open louvers can create glare in open plan offices whenever



Mirror blind in Retro function

Functional principle:
Mirrors to redirect daylight

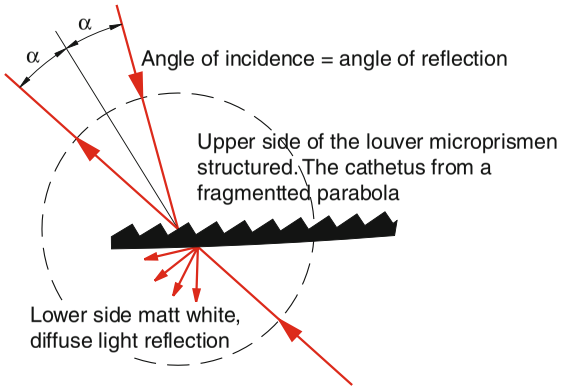
The overheating high summer sun can only be retroreflected with closed blinds.

No visual transmission, darkening effect, electric lighting needed

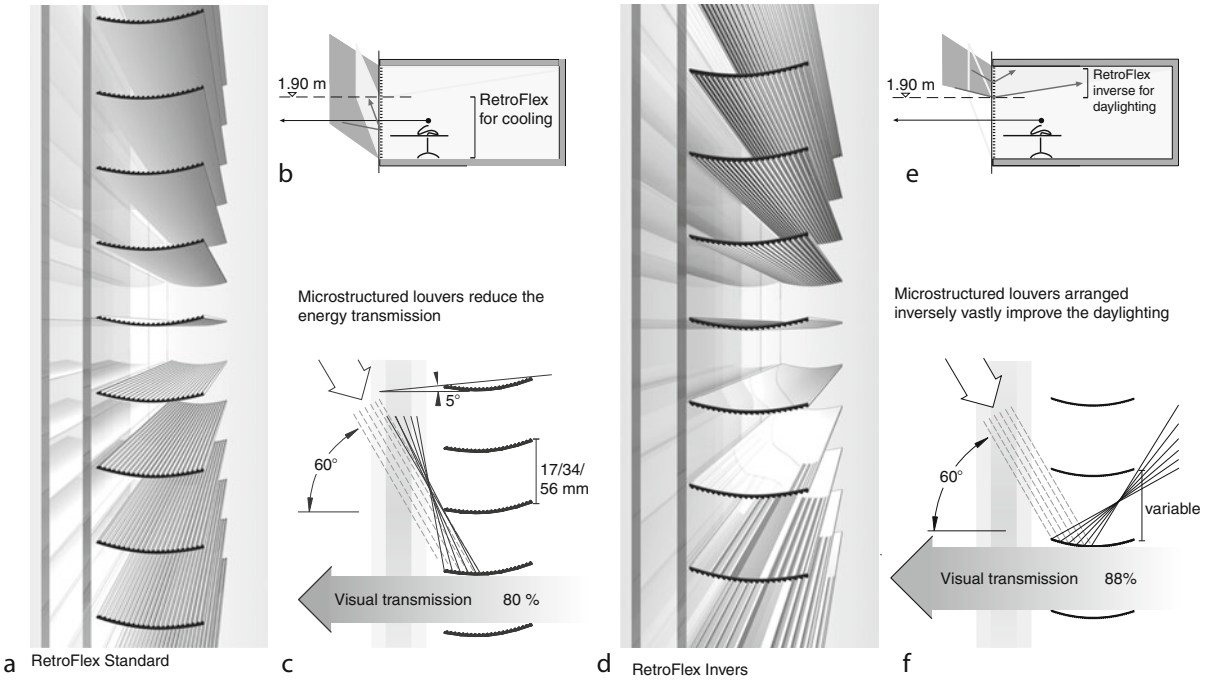
Problem:
Protection from the overheating summer sun is only achieved when the blinds are closed. The redirected solar rays are partially absorbed and transformed into heat on the lower side of the blinds.

Solution:
The upper side of the louver is microstructured, following the logic of a fresnel mirror.

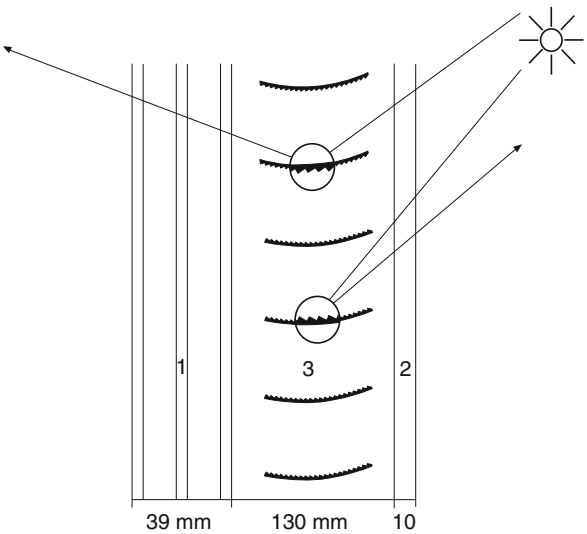
Advantage:
Blinds stay open, visual and daylight transmission is achieved



Daylighting Controls, Performance and Global Impacts. Figure 17
Genesis of a daylight reflecting louver [US 6,367,937; US 6,845,805; EP 1212508]



Daylighting Controls, Performance and Global Impacts. Figure 18
Louvers with microstructured upper surfaces to reflect solar heat gain back outside while supporting diffuse daylighting and views and micro-structured undersides to redirect daylighting into the depth of the room [US 6,367,937; US 6,845,805; EP 1212508]

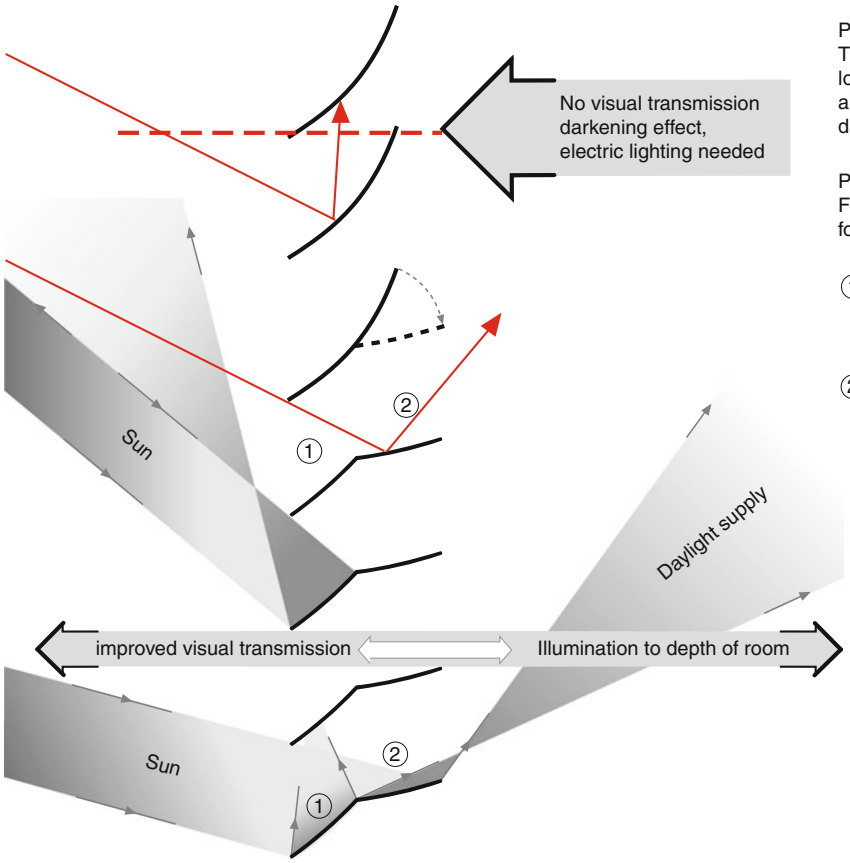


Daylighting Controls, Performance and Global Impacts. Figure 19
Non-ventilated double skin element façade with RetroFlex. U-value of façade 0.65 W/m^{2}K, SHGC value of glazing 0.05–0.1 with open louvers, depending on angle of incidence

the sun angle is low, streaming between the blinds into the space.

In addition, studies of daylight redirection blinds revealed a second area of concern. With higher sun angles in the summer, or a greater steepness for the blind, some of the daylight that is reflected off the blind will fall onto the bottom side of the louver above. The underside of the louvers will absorb a portion of this reflected light and heat up the window zone, even in summer (Fig. 17).

These studies led to the development of more optimized light redirecting louvers that combine surfaces and geometry to manage light in different seasons. One strategy would be to pursue micro-structural or prism innovations, embossing fresnel mirrors onto a concave louver (RetroFlex type louver systems) (Figs. 18a, c, 19, 31, 36). With a crafted fresnel surface, an open, horizontal louver will reflect the sun back to the outside to prevent overheating while allowing for 80% views and diffuse daylighting.



Problem with traditional mirror blinds:
The sun is redirected onto the stone grey lower side of the upper blind and is absorbed. The heat is trapped inside and daylight autonomy is lost.

Problem solutions:
Further development through centeredl folding – opening of the blinds

① 1 Section retroreflecting irradiated energy = passive cooling (protection from overheating)

② 2 Section redirecting daylight = improved daylight autonomy (daylight supply function)

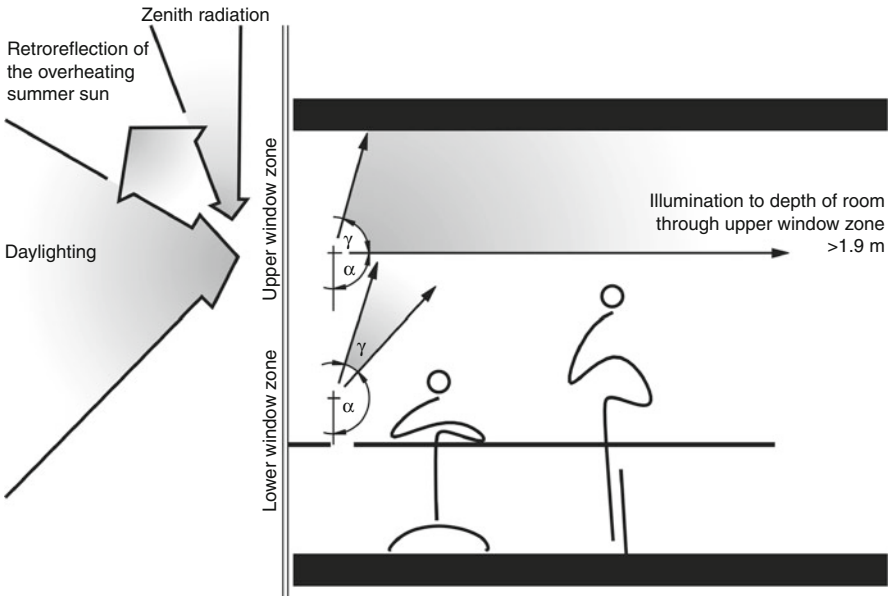
Daylighting Controls, Performance and Global Impacts. Figure 20
Folded mirror louvers as exterior blinds with integrated protection and supply functions through formation of two functional mirrors, a retroreflector toward the outside and a light shelf toward the inside [DE 10260711]

Above the seated window, the mirror prisms can be placed on the underside of the concave blind so that light falling onto their reflecting upper side is redirected to a greater degree into the depth of the room (Figs. 18d, f, 37).

The combination of sunlight-reflective louvers in the viewing area and underside micro-structured louvers above 1.8 m creates a blind assembly that protects the occupants near the window from overheating and glare, while ensuring more effective daylighting, entirely without the loss of views. Since all louvers are positioned identically depending on the sun's angle of incidence, they are easy to operate.

These micro-structured louvers are less than 0.4 mm thick, have a width between 25 and 80 mm depending on density desired, and be positioned inside or in-between a double skin façade (Photos 16, 17, Fig. 31).

While micro-structured louvers offer significant gains over conventional blinds, the best approaches to minimizing heat gain in summer while maximizing daylight requires more challenging geometric solutions. A critical first step is to divide each louver into two sections – redefining external and internal edges with different optical characteristics to respond to the different seasonal positions of the sun in the sky. Given the distance of the louvers from each other, the summer sun only falls onto the section of the louver oriented toward the outside. The outer half of the louver can then reflect the high, direct sun back out for natural



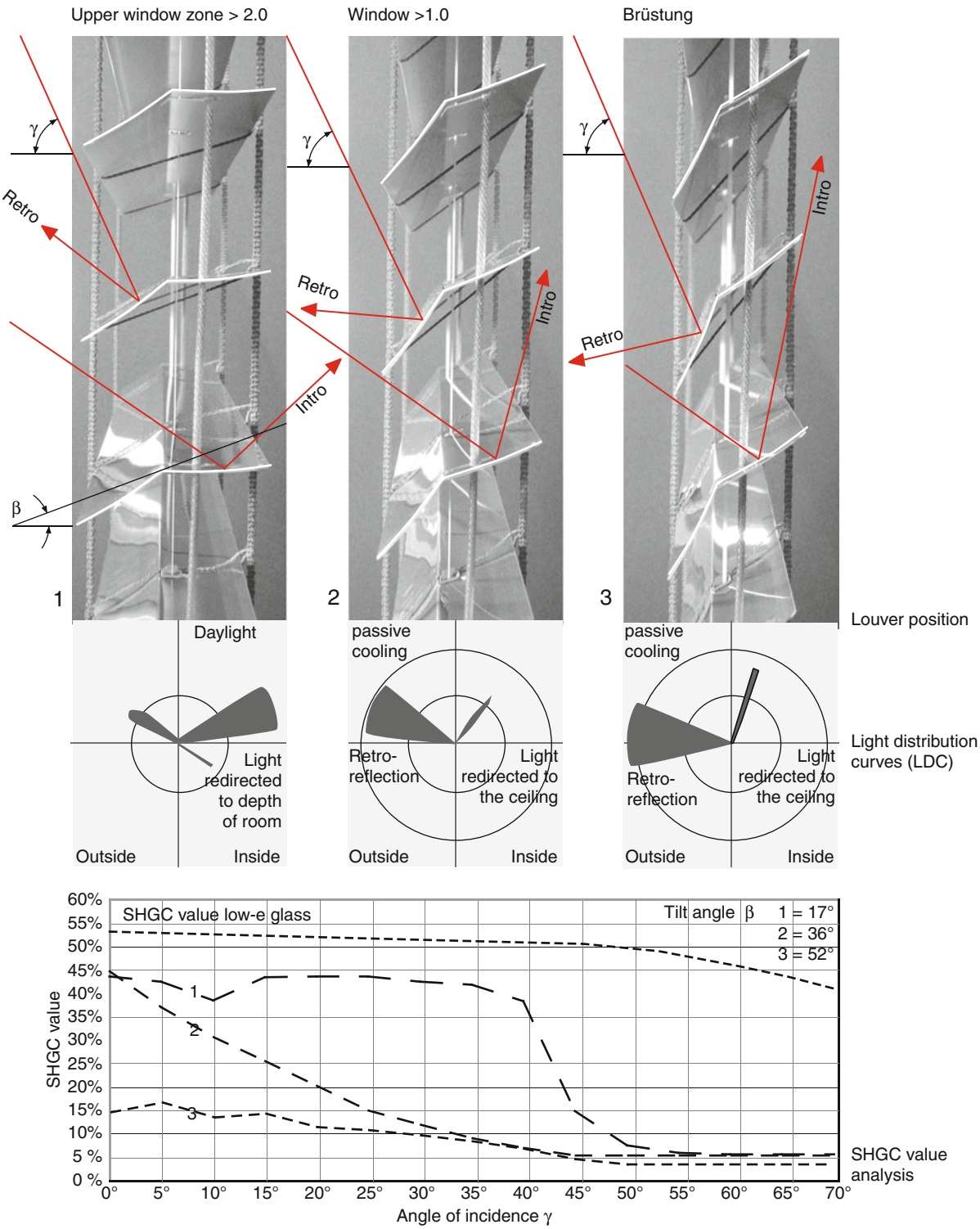
Daylighting Controls, Performance and Global Impacts. Figure 21
Angles of light redirection façade

cooling, while the inner half of the blind can direct the low-angled sun inside to improve illumination of the interior and support passive heating (Fig. 20).

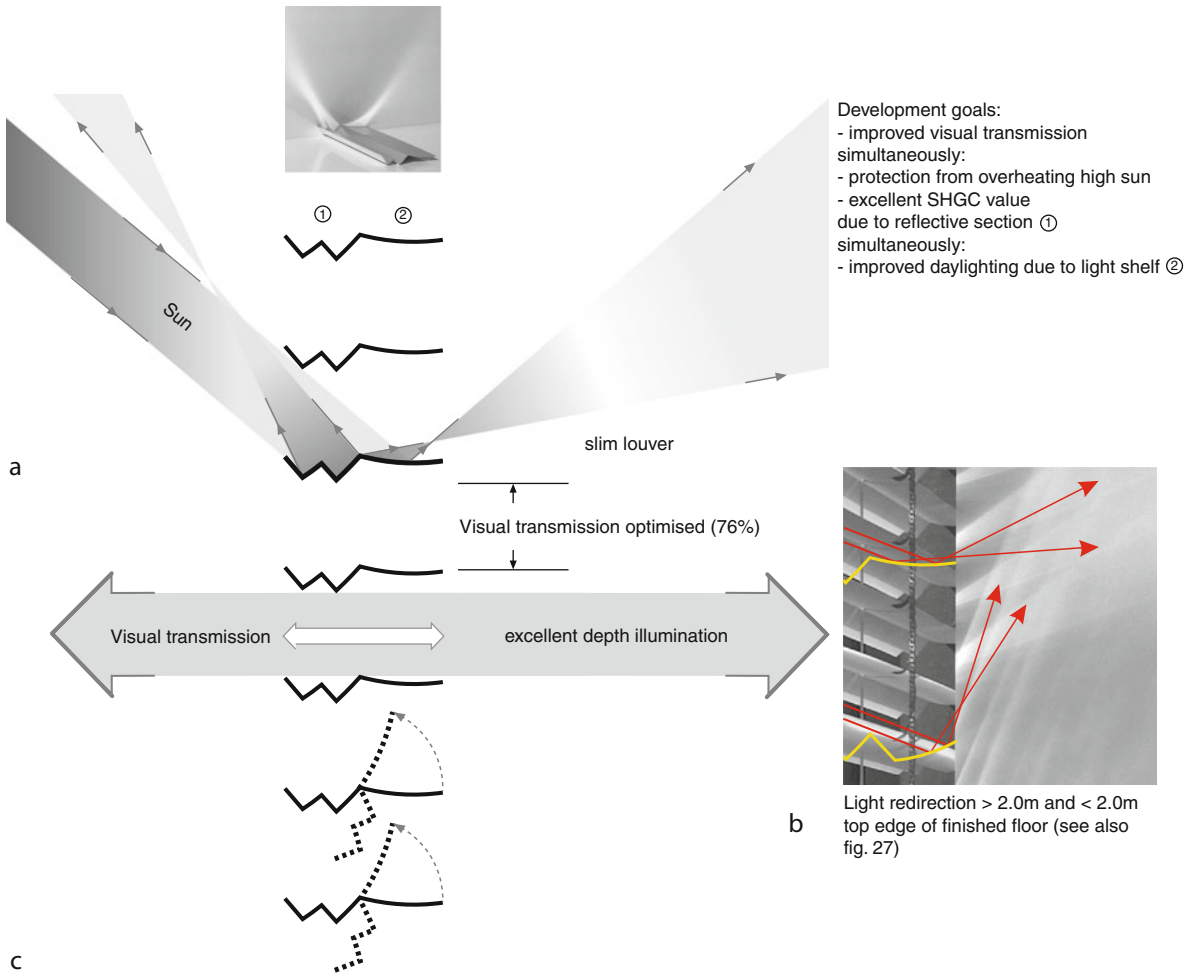
Daylight-optimized louver geometries thus ensure solar heat control based on the sun's angle of incidence without the need for continuous readjustment of the louvers in line with the incident sunlight. On the outer section of the louver, the high summer sun is monoreflectively redirected back into the sky (protective function) and on the inner section of the louver, low incident sun in winter is directed inside (supply function) (Figs. 15, 20, 22, 23). The blinds only need to be closed when the sun is very low and grazing light between the louvers causes glare. The geometries of these blinds and their degree of separation are established based on climate, latitude, and orientation, but there are some truisms for the designer. The further the building is from the equator, the lower the sun is and the longer the heating season is. The angle of incidence increases the closer you get to the equator, along with the temperatures.

The height of the work plane and the eye position of both seated and standing building occupants is an important consideration in daylight design for both glare control and quality views. While the use of reflected light off of the ceiling plane provides good uniformity over a greater depth of the space, any blinds below typical sill height of 1 m or less may create glare by reflecting light directly into the eyes of the occupants. To optimize the illumination into the depth of the room, it is necessary to split the façade into upper (clerestory), middle (viewing), and lower (below the sill) zones. In the upper, clerestory section of the window, daylight should be redirected at a shallow angle to maximize daylight distribution deep into the room. In the middle, viewing section of the window, a somewhat sharper angle is required to redirect the light toward the ceiling to eliminate direct glare, while still maintaining as clear a view as possible. In the lowest section of the window, below the normal sill line, the sharpest reflection angle is important (Figs. 21, 22, 15, 23) (Photo 18).

In addition to the V-shaped geometries described, a W geometry can also support effective daylighting and views with reduced glare and overheating. The W-shaped louver for interior use can offer a slimmer profile in manufacturing and support 76% visual transmission in an open position while the high summer sun is simultaneously reflected outside (Fig. 14).



Daylighting Controls, Performance and Global Impacts. Figure 22
Three Louver positions in floor-to-ceiling blinds for optimum optical performance; analysis of the SHGC values of blind sections 1–3 are an average for all angles of incidence [DE 10260711]



Daylighting Controls, Performance and Global Impacts. Figure 23
Functions of the macrostructured louver "RetroLux" in the upper and lower parts of the blind for glare-free redirection of daylight into the depth of the room and toward the ceiling [US 6,240,999; US 6,845,805; EP 0793761; EP 1212508]

The term "retro" refers to the deflection of light, i.e., the protective function of systems through which the sun is reflected back into the sky. The result, however, can be unwanted glare in the interior due to reflections from the glass. The skill in developing "retroreflective" systems is to guide the light in such a way that the sun rays do not glare in the glass (Photo 27a/b) and thus not visible from the seated position. The light redirection system must be designed such that the sun reflected in the glass hits the bottom side of an upper louver instead of reflecting inside. Louvers following this design rule ensure glare-free transparency of the glass. At a high solar elevation, the visual transmission of such blinds is between 70% and 80% [3].

Macrostructured louvers can also be designed as an integral component of insulated glazing. At 20 mm in width (Fig. 24; Photos 20, 21, 22), the louver section is designed in the shape of a V to stabilize the louver and keep it from bending (Photo 15).

Light-redirecting louvers can also be installed in glazed roofs and set at different angles. Here, the louvers should preferably be installed in a fixed, pre-calculated position in between the insulation glass.

Daylighting Controls, Performance and Global Impacts. Photo 18
Lighting requirements of a daylight façade for glare-free workplaces

Daylighting Controls, Performance and Global Impacts. Photo 19
Daylight façade with RetroLux (Fig. 15), municipal works, Bochum, Germany. Architects: Gattermann + Schossig, Cologne

Versions which can be tilted and tracked like blinds are, however, also possible (Photo 23).

Designing Daylighting: Controls – Balancing Light, Shade, and View

Managing glare, views, and shading as well as electric lighting and air-conditioning energy savings is strongly dependent on louver adjustability and control. On the one hand, controls should respond to weather changes, while on the other hand, changes should be kept to a minimum to avoid annoying users.

While energy optimization might recommend continuous blind adjustment, building occupants might prefer no adjustments, suggesting that louver controls be kept to a minimum of one to three positions throughout the day. East and west façades certainly will need the louvers to be adjusted when the sun rises and sets, and personal preferences will also suggest that some level of control be provided.

Cloudy skies specifically can cause significant glare due to sky brightness. Lights are often switched on during the daytime to avoid contrast glare between the bright sky and the unlit window frames or to compensate for lower illuminance further in the room. Sunny skies can result in direct glare whenever sun angles are low enough to enter directly, and blinds

Daylighting Controls, Performance and Global Impacts. Photo 20
RetroLux Therm in insulation glass (Fig. 24)

are often closed as a result. Some diffusing shades, designed to reduce direct glare and brightness contrast while allowing views, actually make the glare even worse, which also results in the turning on of electric lights. The key to effective daylighting is the ability to dim and redirect light and thermal energy from solar irradiation – independently.

Blinds can be controlled manually with simple guidelines, or controllers can be programmed to respond to calendar or light sensor information to adjust blinds for the optimum balance of view, daylight, and solar heat. Well-detailed blinds can ensure quality views, effective daylighting, at the same time as effective shading, with controls simply extending the long-term performance. Even with the blinds in a permanent down but open position, views can be excellent, while the carefully designed blind geometries and surfaces ensure seasonally differentiated daylight redistribution to the inside and solar heat reflection back to the outside (see Photos 24, 25).

Additional energy savings can then be provided through controls. During the summer, adjusting the light redirection louvers into a closed position when the sun is low on the east or west facing sides of the building can significantly reduce solar energy transmission. Closing the blinds at the end of each work day and opening them again in the morning before work begins can provide additional shade in the summer, or protection against heat loss in winter, during times when daylight and view are not needed (Fig. 25). It is, however, important to open the blinds during the day in order to gain sufficient daylight and views and save electric lighting energy.

In conclusion, the time of day and seasonal dimming and redirection of daylight can best be achieved by careful design of horizontal blinds – their geometries, surface qualities, and levels of adjustability (Photo 26, Fig. 17).

Designing Daylighting: Light Redirection Ceilings

The ceiling is an integral component in effective daylight distribution. Well-designed electric fixtures capitalize on the quality of the reflector as well as the effectiveness of the lamp and lens. Large windows without light redirection systems result in an uneven illumination of the room with excessively high illuminances in the window zone and considerable brightness contrasts (Photos 27a, b). Daylight redirection louvers shift the focal point of daylight design to the ceiling (Fig. 26).

Daylight reflected from the window to the ceiling allows light to be redirected onto the work surface with similar effects as those of conventional electric lighting, but without shadowing (Photos 29, 30).

Reflecting ceilings bear the risk of glare, particularly when viewed from deeper in the room. Consequently, light redirection ceilings and their reflective characteristics and geometry need to be carefully planned. At a very minimum, the ceilings must have the maximum possible light reflection properties, with a modest level of diffusion to reduce glare.

At an optimum, the design requirements for a light direction ceiling equal those of parabolic surfaces

Daylighting Controls, Performance and Global Impacts. Photo 21
The glazed facade is retro-reflecting the overheating sun but supplies the room with more daylight via the ceiling. The workplace is glare-free illuminated. In the openable window area the blinds are moved up to show the effect of the sun without redirecting louvers

Daylighting Controls, Performance and Global Impacts. Photo 22
Examples of illuminating a room using RetroLuxTherm in insulation glass as shown in Fig. 24 (Photo: Helmut Köster) [21]. The redirecting louvers are integrated as fixed systems within the insulation glass of the large glazing. Interior, behind the openable windows venetian blinds with redirecting RetroLux louvers are mounted

Daylighting Controls, Performance and Global Impacts. Photo 23
Light deflection in roof glazing either in insulation glazing or below the glass roof, e.g., with RetroFlex louvers

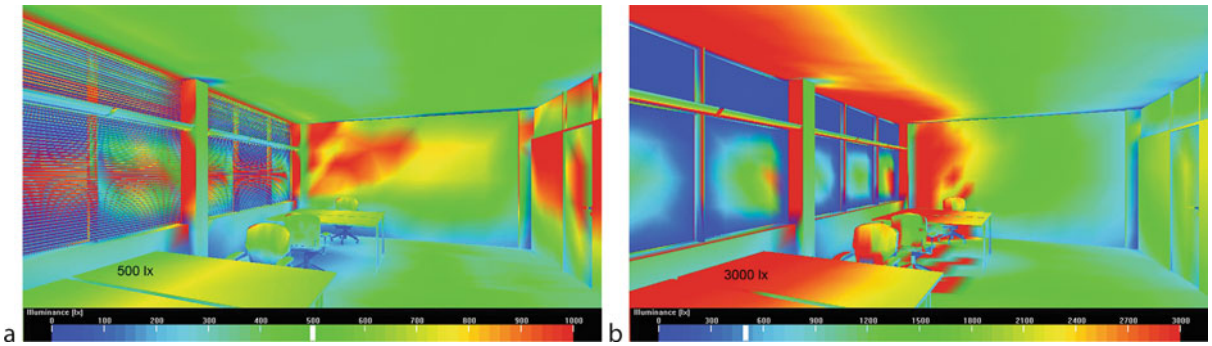
within luminaires. The light redirection and the cut-off angles of the ceiling must be precisely defined and are subject to the requirements of DIN EN 12 464. Superior redirection characteristics can be achieved by using micro-prism structures which redirect the light from the window downward in precise angles due to their angled prism edges. Prism-structured, convex-shaped louvers can be arranged parallel to the façade in order to ensure sufficient dispersion while reacting to different angles of incidence of the daylight.

Daylight reflective ceilings serve the purpose not only of redirecting daylight, but also of redirecting complementary electric lighting that may be installed at the façade, as described in the next section (Fig. 27).

Daylighting Controls, Performance and Global Impacts. Photo 24
RetroLux in solar protection position and closed position

Daylighting Controls, Performance and Global Impacts. Photo 25
RetroLux in solar protection position and closed position

Daylighting Controls, Performance and Global Impacts. Photo 26
View out of RetroFlex blinds in active sun protection position for $\beta = 13$

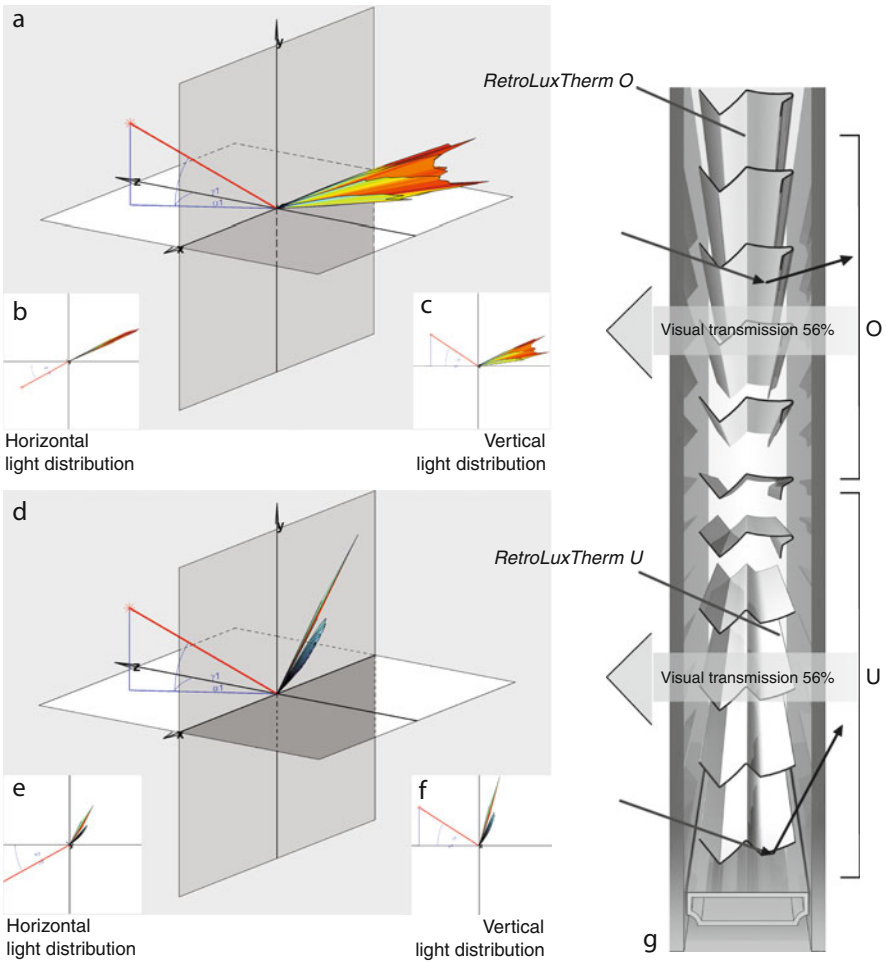


Daylighting Controls, Performance and Global Impacts. Photo 27
(**a**) Light irradiation as a false-color rendering – without redirected light: excessive overexposure (**b**) with redirected light: indirect daylighting (dimmed to comfortable level)

Integration of Daylight and Electric Lighting

European workplace regulations [1] often stipulate individual access to daylight and visual connection with the outside, such that permanent workplaces are frequently found within 6–7 m of windows. However, deep open-plan landscapes are common in many other countries, with boxed-in workplaces separated by seated or standing height partitions that eliminate effective daylight and visual access to the outdoors, and consequently necessitate daytime use of electric lighting.

In Europe, workplaces which do not provide a seated visual connection with the outside are considered inferior. In the USA, no laws exist to ensure seated views of daylight for workers, resulting in the extensive use of electric light during the daytime. In emerging hot countries as well, offices are being moved away from the window to reduce sunlight exposure. While on the surface, rooms without daylight exposure are easier to cool, the electricity used for lighting often exceeds that used for cooling. It is here, in particular, that daylight façades capable of managing solar factors and heat loads, climate by climate, will have the greatest benefits.

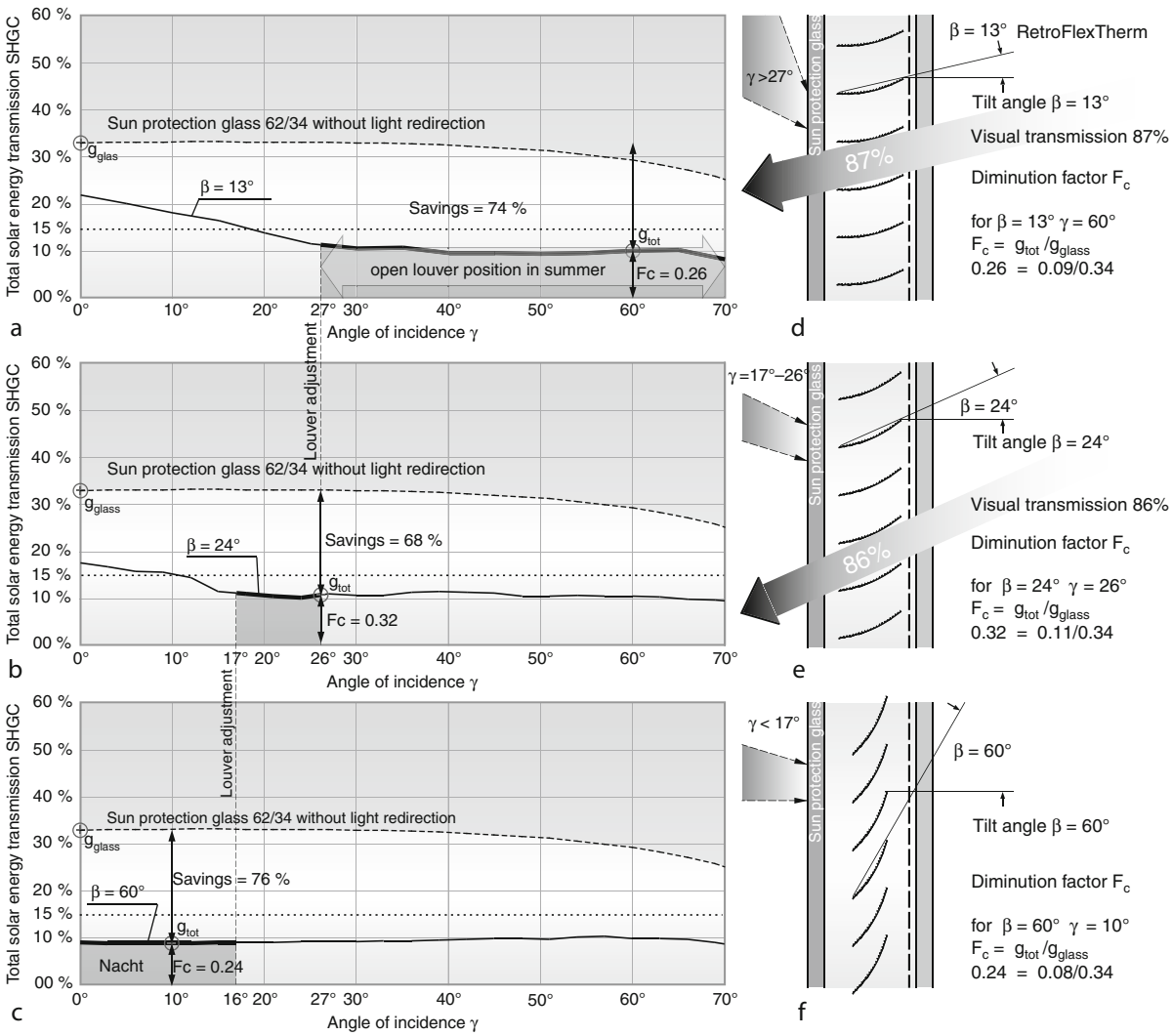


Spatial distribution of the luminance coefficient for $\alpha = 30°$ $\gamma = 30°$
Radiometrically measured values TU Berlin, Institute of Lighting Technology, Prof. Dr. Kaase / Dr. Aydinli

Daylighting Controls, Performance and Global Impacts. Figure 24
Macrostructured louver "RetroLux Therm" in insulation glazing. Analysis of light distribution to the inside for illumination into the depth of the room (type O) (Fig. 24a) and toward the ceiling in the lower window section (type U) (Fig. 24d)

First façades and interior layouts must be designed for effective daylighting, as previously discussed. Then, electric lighting must be designed to complement the daylight, filling in when and where daylight is inadequate.

Traditionally, lighting design considers buildings exclusively at night. To date, the dogma of uniform interior illumination has largely been the priority and a feature of good lighting design. The demands for uniformity may be one reason why daylighting has been so quickly abandoned, in addition to the unmanaged solar heat gain that often accompanies daylight. Workplaces in the window zone will always be brighter than those further inside. The uniformity of illumination can be improved by redirecting the light from the window via the white, reflective ceiling into the depth of the room. Diminishing daylight should then be complemented by increasing levels of electric light, shifting the lighting design goals from total uniformity to ensuring task lighting levels and managing brightness contrast throughout the workspace.



Daylighting Controls, Performance and Global Impacts. Figure 25
SHGC value analysis of a RetroFlex type blind (Fig. 25d–f) with microstructured louvers (fresnel optics) adjusted in two positions

Electric lighting can then be controlled based on the illumination levels needed at various workstations in the room. This is effected either through dimming of individual lamps as needed, or through cascade switching, where the lights furthest from the window are turned on first, cascading on row by row as daylight diminishes. Daylight-responsive controls for electric lighting are minimum requirements in energy-optimized building.

However, lights should remain switched off during the daytime, at least within the first 6 m of the window, to maximize use of the natural, free resource of daylight. Recently, a new generation of "integral" controls have become available which switch on the electric lighting whenever blinds are closed. These types of control are the result of observations that conventional interior or exterior blinds darken buildings even when the sun is shining outside! The promise of daylight technology cannot be fulfilled with such cross-purposes.

As a minimum requirement, daylight autonomy should be at least 80% for the first 6 m into the

Daylighting Controls, Performance and Global Impacts. Figure 26
RetroTop louvers have a microprism-structured surface. Used as a light ceiling parallel to the façade, they redirect both the electric light and daylight. The microstructuring of the RetroTop ceiling also provides for superior acoustic baffling boosted by louvers mounted in a freely oscillating system. By providing the louvers with water-filled tubes, the ceiling becomes a highly effective cooling system whose surface, enlarged through microstructuring, has a cooling effect as it absorbs heat

Daylighting Controls, Performance and Global Impacts. Photo 28
Light redirected through RetroFlex-blinds. To prevent overheating and overillumination only 2–3% of the solar irradiance is redirected onto the ceiling creating 500 lx in 7 m of room depth

room (Fig. 29). In hot climates, effective daylight should be provided without exceeding a maximum total solar energy transmission (heat load) of 10–12%. Indeed, good visual transmission (views) and optimized daylight conditions in the interior can be achieved at even 5% total solar heat energy transmission through the façade. In cold climates, heat gain with light gain will be paired goals, and geometrically optimized and controllable blind design can effectively manage climates that have both heating and cooling loads.

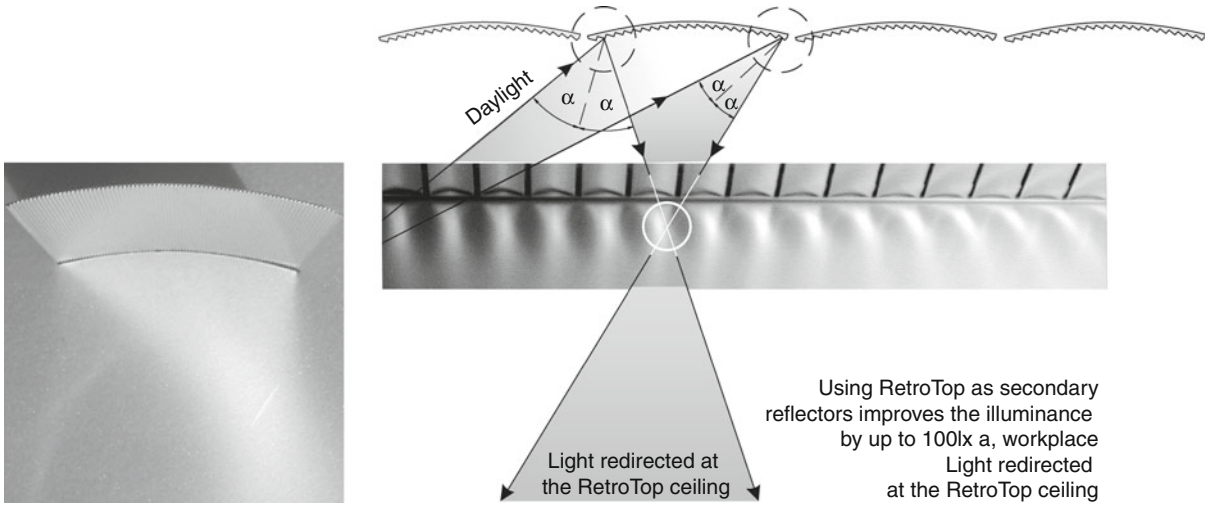
Daylighting Controls, Performance and Global Impacts. Photo 29
The light deflected from the façade to the ceiling is redirected in a conical shape onto the workplace through the concave design of the louvers. This ensures excellent illumination of the workplaces with glare-free top and side light

Daylighting Controls, Performance and Global Impacts. Photo 30
The light deflected from the façade to the ceiling is redirected in a conical shape onto the workplace through the concave design of the louvers. This ensures excellent illumination of the workplaces with glare-free top and side light

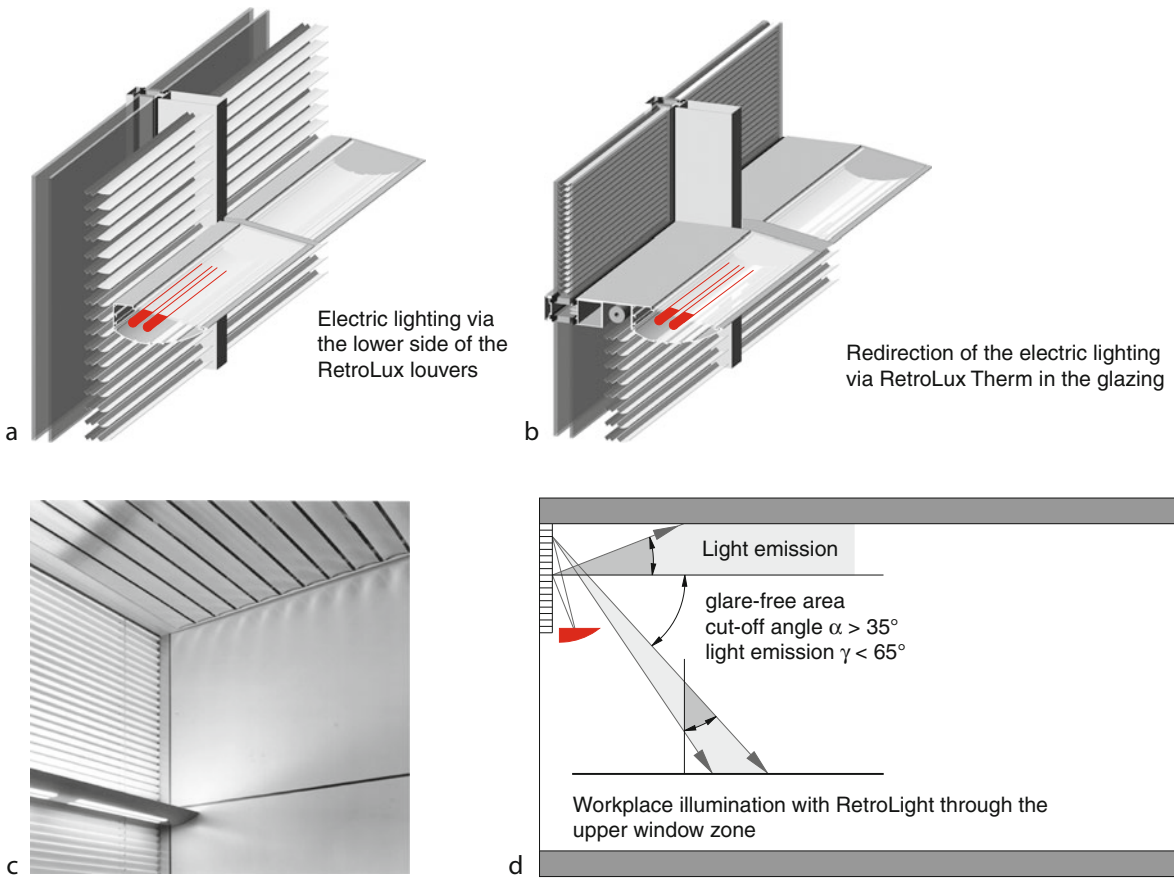
The aim of intelligent combined daylight–electric light controls must be:

- To modulate blinds to provide sufficient daylight for at least the first 6 m, while effectively shading and eliminating glare
- To modulate blinds to reduce the luminance of the window to a comfortable level to ensure both quality views and daylight transmission
- To modulate blinds to reflect out direct solar energy through mirroring in summer, while simultaneously redirecting a percent of the diffuse sky into the interior
- To modulate blinds to reflect in direct solar energy through mirroring in winter, ideally to thermally absorptive surfaces
- To infill electric light when and where daylight is inadequate for the task

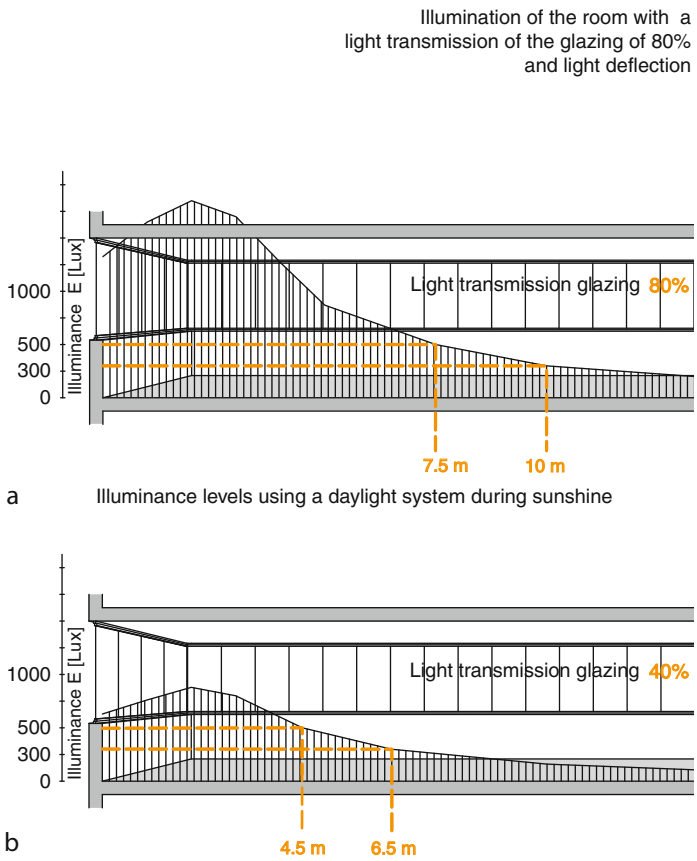
To optimize the integration of daylight and electric lighting controls, at least one major source of electric lighting should shine from the window into the room. This is perfectly possible by integrating the luminous flux and arranging an asymmetrical luminaire with indirect lighting at the height of the window header (Photo 28). This ensures that the bottom sides of the

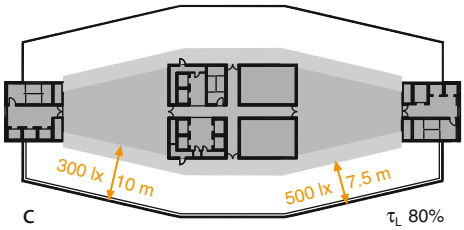


Daylighting Controls, Performance and Global Impacts. Figure 27
Light redirected by ceiling elements. Ceiling louvers with microprism-structured mirrors deflect the indirect daylight redirected toward the ceiling vertically onto the workplace



Daylighting Controls, Performance and Global Impacts. Figure 28
Electric lighting and daylight system optically integrated [EP 0461137; US P 5293305]





Quality assurance

Best practice facades ensure simultaneity of the following quality aspects

- visual transmission > 70%, simultaneously
- diminution of energy irradiation to SHGC< 5–15%, simultaneously
- illumination into depth of room

Daylighting is also a function of the light transmission factor of the glazing.

Potential annual savings in cooling and electric lighting of up to €14.00/sqm of glass facade at €0.15/kW for electricity.

Daylight technology reduces the need for cooling by up to 60% compared to coloured interior blinds. At the same time, the light transmission improves by up to 100%.

Daylighting Controls, Performance and Global Impacts. Figure 29
Comparison of illumination into the depth of room using a daylight redirection system for glazing with a light transmission of 40% and 80%, respectively

daylight redirection louvers are also used to redirect the electric lighting. The resulting integration of redirected daylight and electric lighting is symbiotic (Photo 31).

Quality Assurance in Daylighting Design

Both simulation and measurement techniques are critical for ensuring the highest level of daylight quality, solar heat management and views. Potential sources of error to be avoided include:

– Inappropriate louver geometry with required daylight reflected externally, without depth internally, or between the louvers themselves resulting in a rise in temperature due to absorption
– Inappropriate reflection characteristics of the surfaces with excessive absorption of heat and light
– Inappropriate louver control for time of day and seasonal variations
– Inappropriate electric lighting control to maximize the use of daylight
– Excessively dark glazing

Having looked at light redirection systems from a functional point of view, the question now arises as to their quantitative effectiveness as thermal and light management systems.

For a quantitative analysis of energy flows for heat and light it is important also to consider the totality of the glazing system including the light redirection system. Depending on the installation position of the light

Daylighting Controls, Performance and Global Impacts. Photo 31
Night Architecture using RetroLight, municipal works, Bochum, Germany, Architects: Gattermann + Schossig, Cologne

redirection system either before, behind, or in between the glass, there are complex, physical processes taking place between the light redirection system and the glazing, including echo reflections in the glass, absorption, heat build-up, etc. Daylighting calculation methods [6] often provide incorrect results, as they fail to consider the light redirection performance of the contoured louvers. For this reason, it is invaluable to physically measure the light and heat transmission of the overall glazing/light redirection system using calorimetric and radiometric measuring methods.

Calorimetric methods measure the electric energy required to keep a heat sink installed behind the window assembly at a constant temperature, and translate the results into the heat transfer of the system (SHGC or g-value). It combines measurement with calculation, given a set of tolerances, which is why measurements using the same window assembly may result in different SHGC values from different institutes.

Radiometric measurements are taken under an artificial sun. These measurements are highly accurate and easy to perform. Parameters measured include the wavelengths of the transmitted irradiation, from which the SHGC, or g-value is calculated (Fig. 30). Radiometric measurements can also determine the light distribution (LDC) of the glazing/light redirection system into the interior for different sun altitude and azimuth angles (Figs. 22, 24, 30).

According to DIN 4108–2, the effectiveness of a sun protection system is described using the diminution factor F_C. The total solar energy transmission of a glazed façade is thus calculated using the following equation:

$$g_{tot} = F_C \times g_{glass}$$

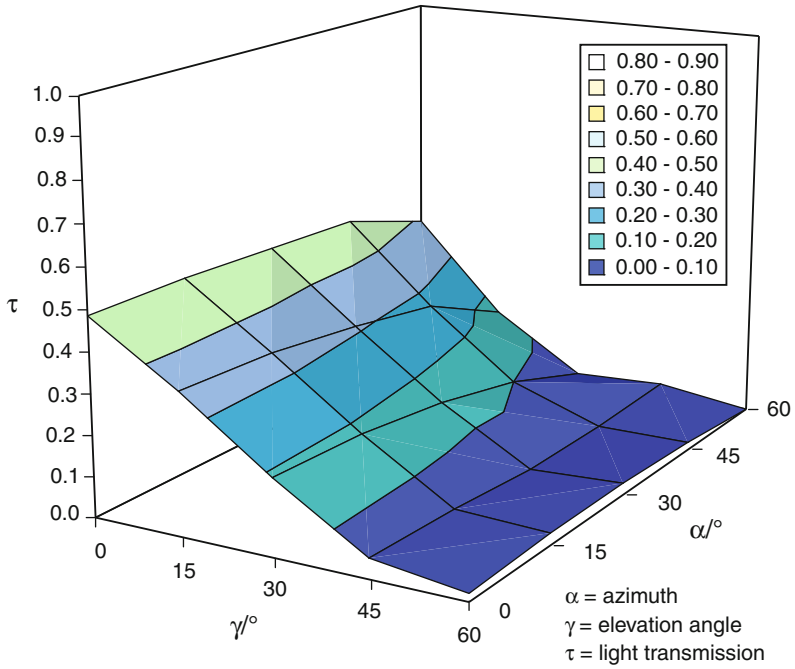
g_{tot} – Total solar energy transmission through glazing including sun protection

g_{glass} – g-value of the glazing

F_C – Diminution factor of the sun protection system

This calculation, however, is not sufficient for quality assurance of the façade design since it is focused on the thermal performance of the window relative to sunshine (Fig. 31). Best thermal performance of the façade will ensure low temperatures of the glazing (Fig. 32). However, windows are critical to effective daylighting, a factor that trumps cooling, so light transmission is equally important for consideration. The balance of solar heat and light is taken into account in the latest DIN EN 13363 standard – yet still without consideration of light-redirecting blinds, especially those that can distinguish high and low sun angle sunshine.

Windows have the additional goal of providing views, visual connections with the landscape and life on the ground plane. While very low solar heat gain factors are easy to achieve with highly reflective glass, or window assemblies that assume shades to be predominantly closed, these windows no longer ensure the views and daylight critical to building occupants (Photos 21, 22, 24, 26, 32). As a result, all



Daylighting Controls, Performance and Global Impacts. Figure 30
Example to illustrate the light transmittance τ for RetroLux Therm O with type 63/32 solar protection glazing (see also Fig. 24)

calculations should assume that blinds are in an adequately open position to ensure views and daylight effectiveness.

To evaluate the quality of the window assembly for thermal and visual performance, the g-value (SHGC) parameters must be complemented by light transmission and view clarity parameters. In other words, the g-value must be differentiated as a combined value of energy transmission for short-wave light radiation and long-wave heat radiation components. In the USA, a light to solar gain ratio (LSGR) has been introduced to give value to daylighting and views in relation to shading in the selection of glazing assemblies.

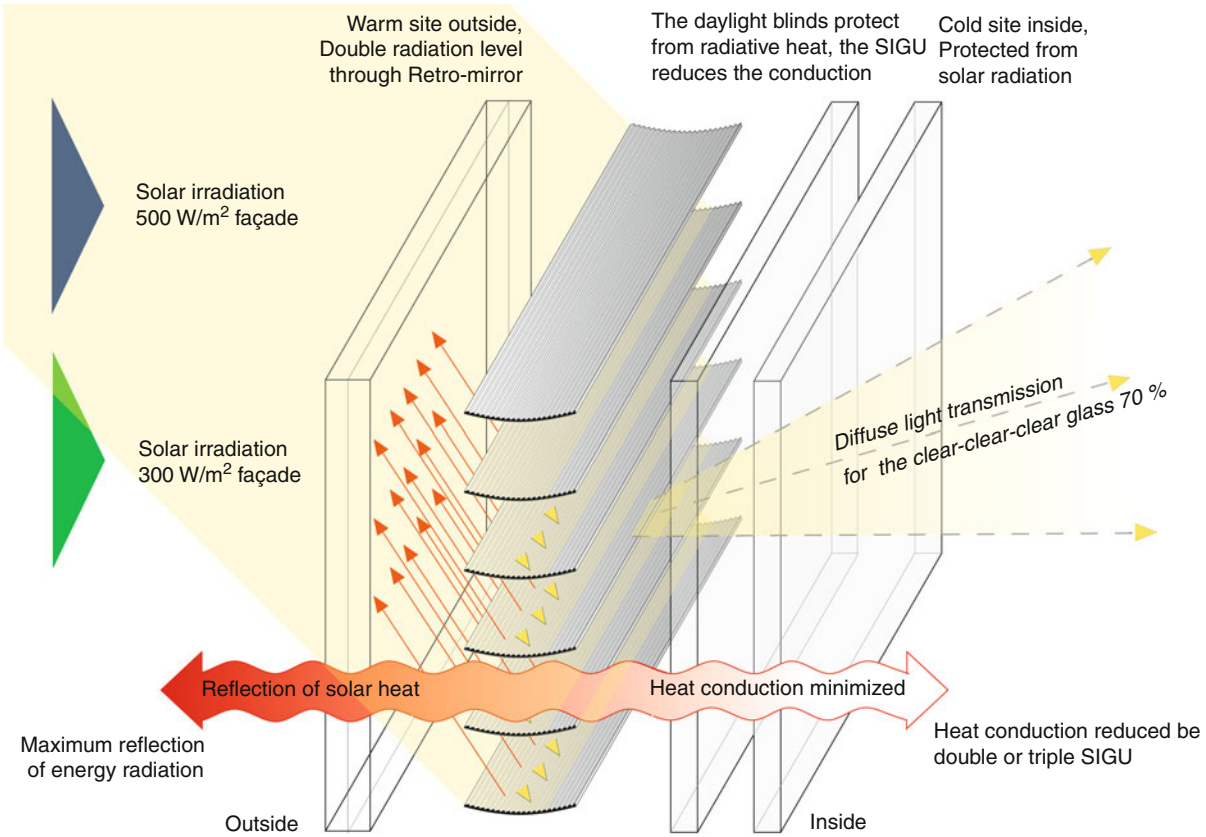
It is important to note that while glazing assemblies may be evaluated by combined values of daylight and view as well as shade, many internal and external layers – shade, louvers, overhangs – are not evaluated for their impact on daylighting. In addition, there are many climate periods when solar heat gain is desirable, such that dynamic properties for the entire window assembly are ideal.

Conclusion

Daylight and Solar Managing Façades as a Focal Point for Building Energy Futures

The largest primary energy load in buildings is for lighting, mostly during the daytime. The most rapidly increasing primary energy load in buildings is for cooling. This is where daylight technology offers the greatest opportunity for the design and engineering community. It not only reduces or manages the external solar load, but also eliminates the electric lighting load during the daytime along with the associated cooling loads.

Integrated design is critical to the detailing of façades for effective daylight and solar energy management. An interdisciplinary process can ensure that it meets a wide range of design goals. To date, it has been common practice in Europe to independently design sun shades and glare protection in different layers inside and outside the building. The design of fixtures

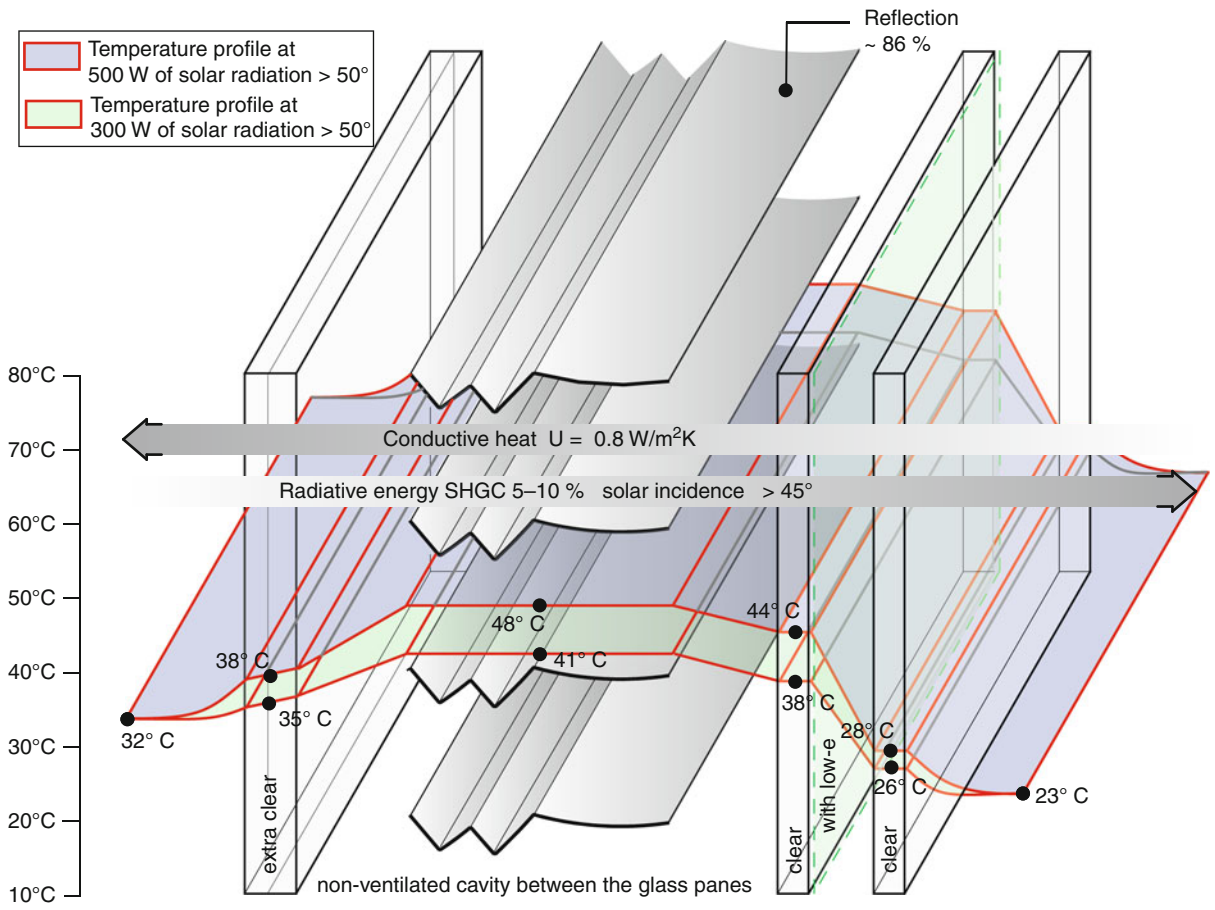


Daylighting Controls, Performance and Global Impacts. Figure 31
Example of optimized façade construction with a closed cavity: the outer pane is exposed to a double radiation load by incident and retroreflected sun. The outer pane should, therefore, be made from low-iron, color-neutral glass to prevent absorption and heat generation. The lower the temperature of the outer glazing is, the lower the heat radiation between the louvers will be on the inner glazing. The better the reflectivity of the light redirecting louvers is, the lower the heat-up of the air space will be

to diffuse light does not consider daylight. Structural design is typically independent of light or thermal design. Interior space layout and ceiling design is independent of daylight and often thermal conditioning. Breaking the building components into individual functional elements frequently led to excessive efforts and significant cost increases in the construction. Integrated design engages the architect, the lighting/daylighting designer, the mechanical engineer, and even the structural engineer in synergistic innovation (Figs. 33–37).

University curriculum must be revised to recognize the importance of integrated design. Every architect should be taught to be effective "system integrators," capable of orchestrating an interactive process that engages the expertise of the mix of disciplines early in the design process for maximizing synergistic innovation. An effective integrator must have a core competency in building physics, climate and energy management, and the full set of building technologies that impact heat, light, air, sound, and structural integrity – with the building façade and adjacent interiors as a primary focus for a sustainable future.

At the graduate level, universities must also take a far more proactive role in research related to daylighting technologies and system integration. The optical



Daylighting Controls, Performance and Global Impacts. Figure 32

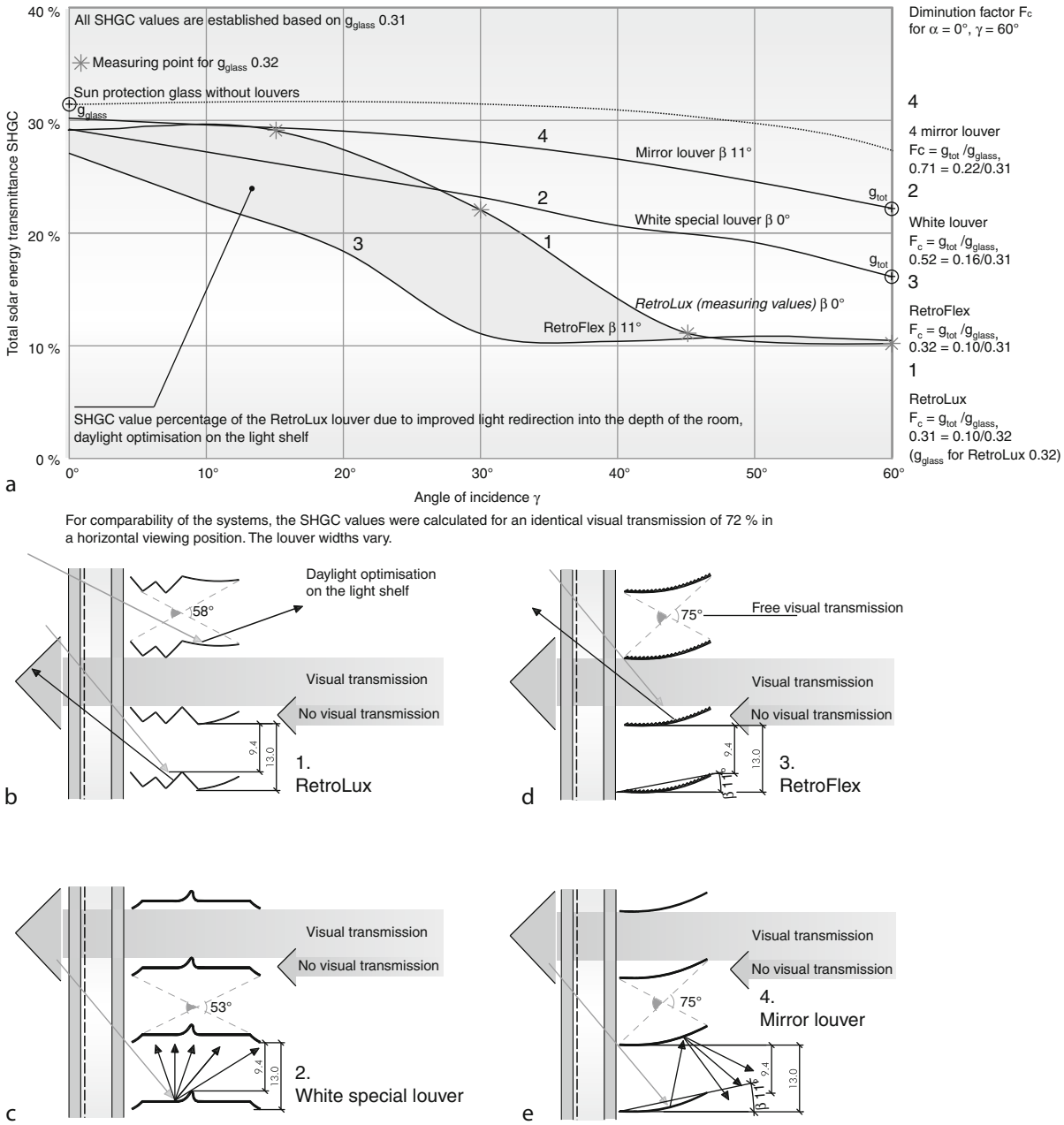
The above temperature profile in a closed cavity can be reached with RetroLux or RetroFlex blinds at angles of incidence >45°. The contour of the blinds is of crucial importance, because the low temperatures can be realized even in a horizontal louver position. Simultaneously a visual transmission of 70–80% between the louvers and the desired improved daylighting is achieved. The temperatures in the cavity should not exceed 60°C to ensure the longevity of the motors, plastic parts, and fibers

research inherent in fixtures that manage point light sources and linear light sources have not yet been applied to planar light sources such as a window. The geometry and positioning of light reflectors (internal, external, between glass installations) and critical dynamics given climate and time of day variations are critical areas of scientific and technical innovation. The interaction between glass coatings, such as reflective enhancing PVD coatings, and mirror systems, such as semi-specular, lacquered or anodized surfaces, offer enormous potential for the research and manufacturing community.

For the design community, algorithms and computational tools need to be developed to calculate the bi-directional energy- and light transmission values for dynamic tilt positions of the special mirror contours with reference to the positions of the sun. Existing simulation tools must be refined beyond qualitative transmission data, to describe the optical properties of reflected light to calculate the effectiveness of

Daylighting Controls, Performance and Global Impacts. Photo 32
Visual transmission of RetroLux

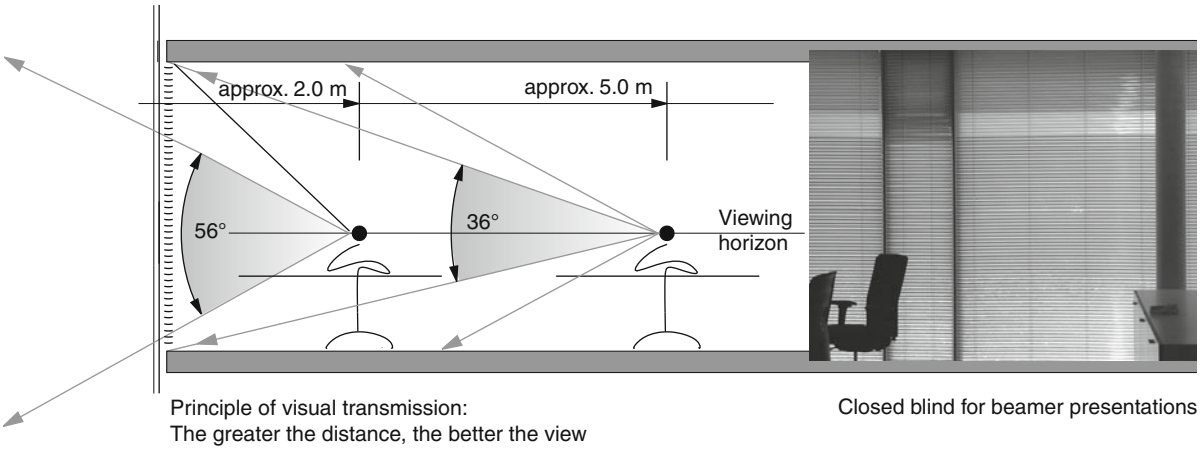
Daylighting Controls, Performance and Global Impacts. Photo 33
The Sopharma and Litex buildings in Sofia have a non-ventilated double-skin façade with a closed cavity (see Fig. 31) and RetroFlex blinds. The room temperature does not exceed 26°C even without chilling even at outside temperatures of 35°C and sunshine



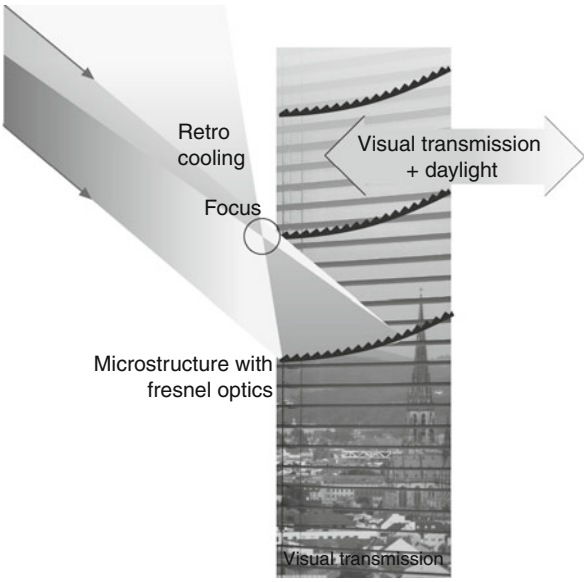
For comparability of the systems, the SHGC values were calculated for an identical visual transmission of 72 % in a horizontal viewing position. The louver widths vary.

Daylighting Controls, Performance and Global Impacts. Figure 33

This illustration shows the dynamic SHGC values of different types of louvers based on condition of the same visual transmission for different angles of incidence. The mirror louver (Fig. 33e) performs very poorly, as the sun is reflected onto the gray lower side where it is absorbed and converted into heat. The white louver (Fig. 33c) shows better thermal properties. The RetroLux louvers (Fig. 33b) with light shelves in the second section produce a good light output with an excellent passive cooling capacity in the high summer sun. Despite open position, the RetroFlex louvers actively provide passive cooling even when the sun is low (Fig. 33d)



Daylighting Controls, Performance and Global Impacts. Figure 34
Analysis of visual transparency for RetroLux and RetroLux Therm during active protection from the high summer sun and closed blinds

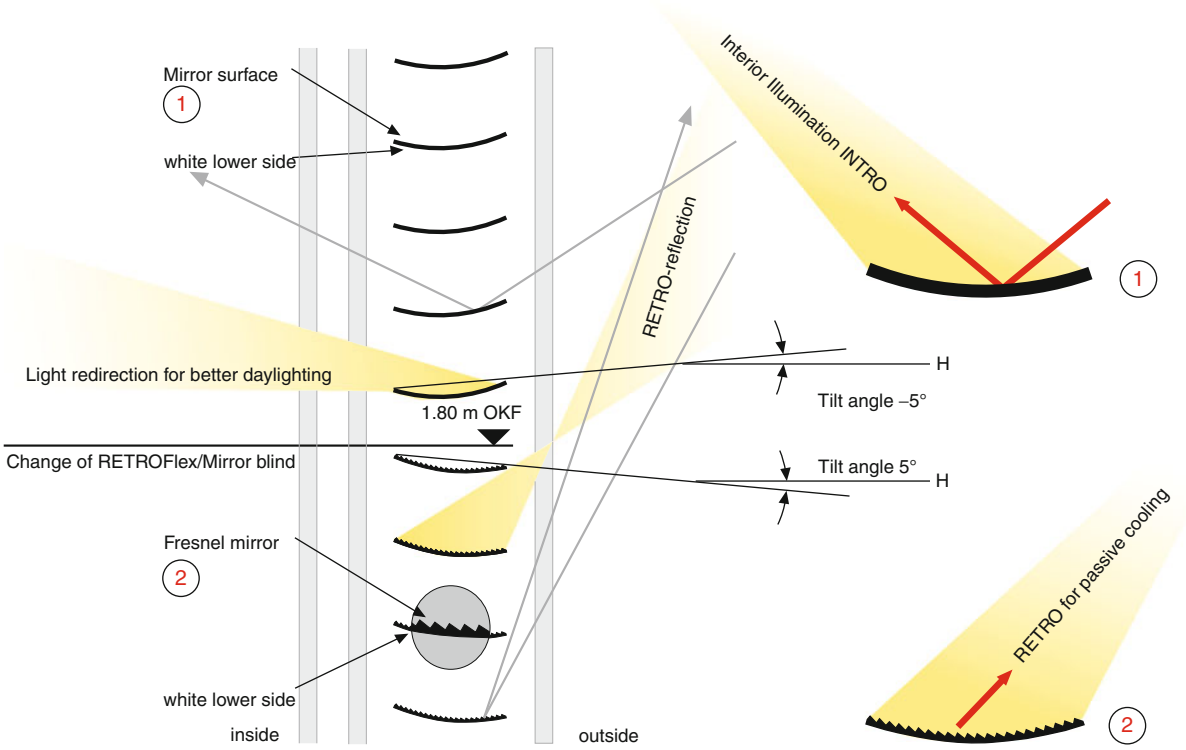


Daylighting Controls, Performance and Global Impacts. Figure 35
Visual transmission through RetroFlex louvers 74–80% (see also Figs. 17, 18)

daylighting for different sky and room conditions. Tools that accurately quantify annual and peak lighting, heating and cooling energy savings, and quantify the carbon benefits will be critical for designers as well as policy makers. Then standards such as DIN EN 13363–1 and 13363–2 [7] need to be refined to reflect the full energetic behavior of façades.

Appendix A. Daylight redirection-systems patents [excerpt]: Dr-Ing. Helmut Köster

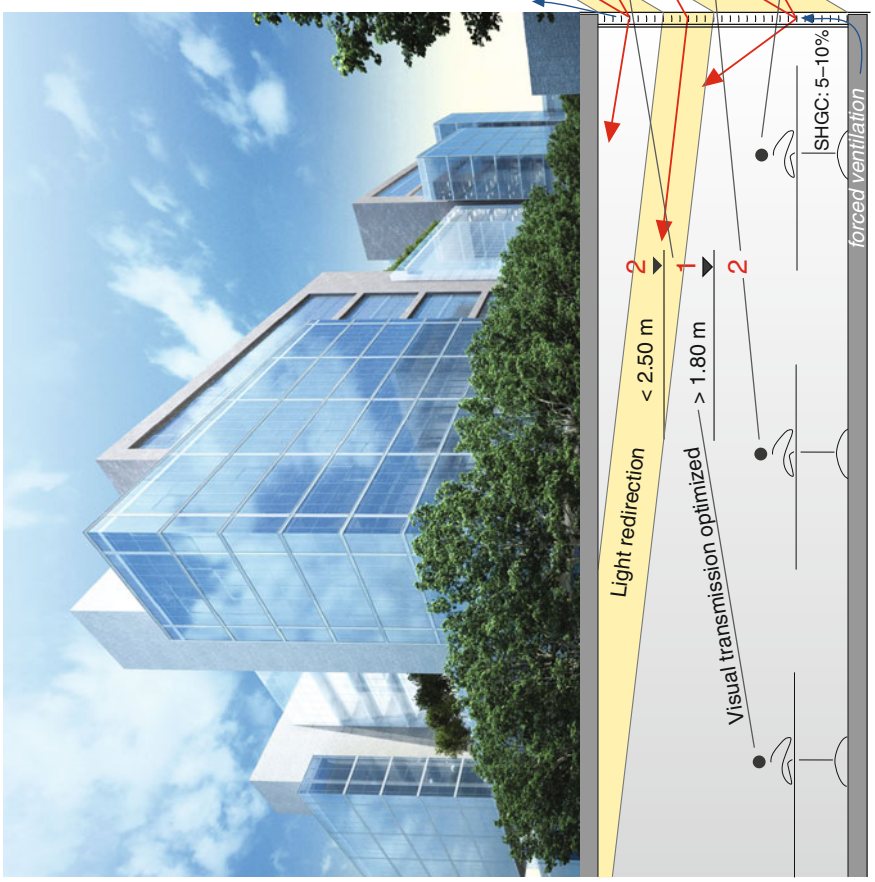
EP 0793 761	Stepped Lamella for guiding Light Radiation
DE P 69514 005.1-08	Stepped Lamella for guiding Light Radiation
CH EP 0 793 761	Stepped Lamella for guiding Light Radiation
IT (EP) 0 793 761	Stepped Lamella for guiding Light Radiation
Fr (EP) 0 793 761	Stepped Lamella for guiding Light Radiation
GB EP 0 793 761	Stepped Lamella for guiding Light Radiation
NL (EP) 0 793 761	Lamelles en Gradins Destinees au Guidage de Rayonnement Lumineux
AT EP E 187 800	Stepped Lamella for guiding Light Radiation



Daylighting Controls, Performance and Global Impacts. Figure 36
Strategy: within the retroreflective RetroFlex curtain a few mirror blinds with opposite qualities may be inserted in the upper window part to illuminate the working area in large depth (see also Fig. 37)

AU P 704 884	Stepped Lamella for guiding Light Radiation
CA P 2 205 560	Stepped Lamella for guiding Light Radiation
US 6,240,999	Stepped Lamella for guiding Light Radiation
EP 0461 137 B1	Lichtlenksystem für die Beleuchtung eines Innenraumes
DE P 590 09 101.8-08	Lichtlenksystem
US P 5 293 305	Light Guidance systems for the illumination of an interior area
USA 6,367,937	Sun Protection Installation.....
AT 411613	Sonnenschutzanlage mit Sonnenschutzlamellen.....

CH 694,947	Sonnenschutzanlage mit Sonnenschutzlamellen.....
NL 1010766	Zonwering met zonweringlamellen.....
GB 2332229	Sun Protection Installation.....
FR 9815482	Sun Protection Installation.....
IT - 1303650	Impianto di Protezione contra....
CA 2,255,302	Sun Protection Installation.....
AU 756628	Sun Protection Installation.....
AU 643429	Light deflecting system.....
DE 100 2006 006 855.6	Bewegliche Fixierung leiterartiger Bauelemente
DE 10 2005 028 6550	Medienfassade
DE 102 60 711	Blendfreie Jalousien



Daylighting Controls, Performance and Global Impacts. Figure 37
In the Standard Bank in Johannesburg, South Africa, the window is structured in three functional sections (see Fig. 36) with opposite values to ensure best daylighting in larger depths of the room without overheating

DE 100 18 451	Herstellung von linearen, prismatischen Strukturen auf einem lamellenartigen Festkörper
DE 198 28 543	Sonnenschutzanlage für Sonnenschutzlamellen, die eine gezahnte Oberseite aufweisen
DE 196 36 817	Sonnenschutzanlage mit Sonnen- schutzlamellen, die eine gezahnte Oberseite aufweisen
DE 44 42 870	Jalousielamelle zur präzisen Steuerung der direkten Sonneneinstrahlung
DE 000 331 483 -0001	Oberfläche für Jalousielamellem _ Oberfläche mirkostrukturiert, gezahnt
M9502488.3	Jalousie zur Tageslichtumlenkung
000 334 483 Alicante	selbst gefertigtes Lamellenprofil Retroflex
DE 401 04 706.7	Fassadenpfostenausbildung vorzugsweise für Glasfassaden mit und ohne Leuchte
DE 401 06 175.2	Lichtlenkdecken (1 Muster)
DE 401 09 455.3	Asymetrisch strahlendes Leuchtenmodul
DE 401 10 472.9	Oberlichtleuchte von Trennwänden
DE 402 02 313.7	Leuchten
DE 402 02 431.1	Trennwandleuchte II
DE 402 03 978	Lichtlenkjalousien bzw. Lichtlenkvorrichtung
DE 402 10 688	Jalousiebehang
DE 403 04 242	Lamellenvorhänge
DE 404 04 133.7	Lamellenvorhänge
M 9502488.3	Jalousie zur Tageslichtumlenkung
DM/052988 (15)	Blinds for reflecting sun and diffuse daylight as well as artificial light (2 x Retroluxtherm)
EP 00951306.0erteilt als EP 1212508	gezahnte Tageslichtlamelle
PCT/EP00/05929Intern. Application No.	gezahnte Tageslichtlamelle, toothed daylight blinds
CA 2,377,711	Toothed Daylight Blinds
USA 6,845,805	Toothed Daylight Blinds
AU 758 794	Toothed Daylight Blinds
EP 2006 005909	Medienfassade
PCT/EP2006/ 005909International Application No.	Medienfassade, s. auch, EP Anmeldenummer 06015154.5

Bibliography

Primary Literature

1. 2008: Verordnung über Arbeitsstätten (Arbeitsstätten-verordnung – ArbStättV), BGBl. I S. 2768
2. Arasteh D, Johnson R, Selkowitz S (1986) Definition and use of a daylight "Coolness" Index. In: International Daylighting Conference, Long Beach, 5–7 Nov 1986
3. ASR 7/1 Sichtverbindung nach außen; ASR 7/3 Künstliche Beleuchtung; ASR 7/4 Sicherheitsbeleuchtung
4. Cakir A, Cakir G (1994) Lixht und Gesundheit. Eine Untersuchung zum Stand der Beleuchtungstechnik in deutschen Büros. ERGONOMIC Institut für Arbeits- und Sozialforschung Forschungsgesellschaft mbH, Berlin
5. Deutscher Wetterdienst, Offenbach (2010) URL: http://www.dwd.de (Stand 09.04.2010). Accessed 9 Dec 2011
6. DIN 4108–2: Wärmeschutz und Energie-Einsparung in Gebäuden. Teil 2: Mindestanforderungen an den Wärmeschutz. Ausgabe: Juli 2003 DIN EN 13363: Sonnenschutzeinrichtungen in Kombination mit Verglasungen – Berechnung der Solarstrahlung und des Lichttransmissionsgrades. Ausgabe September 2007
7. DIN EN 13363–1: 2003 + A1:2007 (2003/2005/2007) Solar protection devices combined with glazing – Calculation of solar and light transmittance – Part 1: Simplified method. DIN EN 13363–2:2005 Solar protection devices combined with glazing – Calculation of total energy transmittance and light transmittance – Part 2: Detailed calculation method
8. Energy Policy Division of the Washington State Energy Office (Hg.): Washington State Energy Use Profile, Commercial Sector. URL: http://www.commerce.wa.gov/energy/archive/FILES/PRFL/docs/aboutweb.htm (Stand: 09.04.2010). Accessed 9 Dec 2011
9. Franke H (ed) (2007) Handbuch für Beleuchtung. Ecomed, Landsberg

10. Further daylight publications
11. Prof. Dr.-Ing. Dr. h.c. Hartkopf V (2009) Transnational endowment for sustainable built environments (TESBE). In: Green Building Conference, Carnegie Mellon University, Pittsburgh, 9 Nov 2009
12. Prof. Dr. Kaase, Heinrich: Ga-Nr.: HK-KÖ2-01 (2001) Gutachten über die Bestimmung der lichttechnischen und strahlungsphysikalischen Kennzahlen von zwei Retro-Lamellen als Innenjalousie. Institut für Elektronik und Lichttechnik der Technischen Universität Berlin
13. Klammt S, Müller H, Neyer A (2009) Advanced daylighting using micro-structured components. TU Dortmund. In: Proceedings of PLDC, Berlin, 27–31 Oct 2009
14. Köster H (2004) Tageslichtdynamische Architektur – Grundlagen, Systeme, Projekte, 1st edn. Birkhäuser, Basel (published in German, English, Chinese and Czech)
15. Patents: see Appendix
16. Dr. Pfafferrott J, Kalz D (2007) Thermoaktive Bauteilsysteme, FIZ Karlsruhe GmbH (Hg.). URL: http://www.bine.info/fileadmin/content/Publikationen/Themen-Infos/I_2007/themen0107internetx.pdf. Accessed 6 Dec 2011
17. Richtlinien VDI (2002) VDI 6011 Blatt 1: Optimierung von Tageslichtnutzung und künstlicher Beleuchtung – Grundlagen. Beuth, Berlin
18. Ruhstorfer W Computergrafik und Bildverarbeitung (1997) URL: http://www.uni-regensburg.de/EDV/Misc/CompGrafik/Script_5.html. Accessed 9 Dec 2011
19. Saint Gobain Glass, Memento (2005)
20. Twarowski M (1962) Sonne und Architektur. Georg D. W Callway, München
21. Umweltbundesamt. Daten zur Umwelt (2011) URL: http://www.umweltbundesamt-daten-zur-umwelt.de/umweltdaten/public/theme.do?nodeIdent=3437 [Stand: 09.04.2010]. Accessed 9 Dec 2011
22. SOLARTRAN Pty Ltd. URL: http://www.solartran.com.au/lasercutpanel.htm. Accessed 9 Dec 2011
23. World Data Center for Meteorology, Asheville. URL: http://www.ncdc.noaa.gov/oa/wdc/index.php?name=climateoftheworld. Accessed 9 Dec 2011

Books and Reviews

Goldmann M (2010) Gelenkte Lichtblicke, Deutsches Architektenblatt, Ausgabe Hessen/Rheinland-Pfalz/Saarland 06/10, S.38–S.40

Köster H (2010) Energiesparressource Tageslicht, Lichtlenkung. industrieBAU 56(2):S.36–S.41

Köster H (2010) Energiesparressource tageslichtlenkende Fassaden. In: Innovative Fassadentechnik, A 61029. Ernst & Sohn, Berlin, S.51–S.56

Köster H (2010) So sind Sie Ihrer Konkurrenz einen Schritt voraus, Diese Tipps von Köster Lichtplanung machen Sie zum Tageslichtexperten. Sicht + Sonnenschutz 9: S.16–S.19

www.koester-lighting-design.de

Desalination Technology for Sustainable Water Resource

Joon Ha Kim[1,2], Miriam Balaban[3]
[1]School of Environmental Science & Engineering, Gwangju Institute of Science & Technology (GIST), Gwangju, South Korea
[2]Sustainable Water Resource Technology Center (SWRTC), Gwangju, South Korea
[3]European Desalination Society (EDS), Tel Aviv, Israel

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Multistage Flash (MSF) Distillation
Multieffect Distillation (MED)
Reverse Osmosis (RO)
Forward Osmosis (FO)
Membrane Distillation (MD)
Future Directions
Bibliography

Glossary

Brine Water that contains a salt concentration greater than 50,000 ppm. Also, it is generally used in MSF processes to describe inlet seawater.

Concentration polarization Higher level of concentration profile of solute nearest to the upstream membrane surface compared with the more or less well-mixed bulk fluid far from the membrane surface.

Forward osmosis Process using osmotic pressure difference between brine and permeate as the driving force for water production.

Fouling The deposition of suspended or dissolved substances on the membrane surface, at the membrane pore, or within membrane pores.

Membrane distillation Thermal-driven membrane process, in which only water vapor can be transported through hydrophobic membranes.

Multieffect distillation Process uses multiple evaporators in a sequence at progressively low pressures. Generated vapor from each effect is used as a heat source in the next effect.

Multistage flash distillation Process using flash evaporation to convert a portion of the inlet seawater into steam in multiple stages and consequently condensate to product water.

Osmotic pressure Pressure causes the movement of solvent molecules through a semipermeable membrane into a region of higher solute concentration, aiming to equalize the solute concentrations on the two sides.

Permeate water Water product that leaves a membrane module and contains penetrants.

Pressure-retarded osmosis Process harvesting energy retrieved from the difference in the salt concentration between higher solute concentration and lower solute concentration solutions.

Reverse osmosis Process requiring application of transmembrane pressure on the liquid-phase pressure-driven separation process that causes the selective movement of solvent against its osmotic pressure difference.

Temperature polarization Temperatures at the membrane surfaces that differ from the bulk temperatures measured in the feed and permeate.

Definition of the Subject and Its Importance

As an alternative solution for eliminating water shortage, desalination processes have drawn an increasing amount of attention. Seawater desalination is a separation process to produce freshwater from saltwater, and can be categorized into two parts: thermal and membrane separation methods. Thermal separation methods are based on principles of the evaporation process and include multistage flash (MSF), multieffect distillation (MED), and mechanical vapor compression. On the other hand, the mechanisms of membrane separation methods are based on solution-diffusion or/and sieving processes. Among these processes, MSF and reverse osmosis (RO) currently cover more than 90% of the total market share. RO processes, in particular, have become competitive with other desalination processes due to their relatively low membrane cost and energy consumption.

Introduction

According to United Nations' reports, around 700 million people in various countries suffer from water shortage problems [1]. In addition, 1.8 billion people

are afflicted with severe water scarcity problems, and it is expected that up to two-thirds of the world population will encounter a number of difficulties from water stress. As such, it is predicted that conflicts may arise between regions where the suffering from water shortages become serious. Moreover, the absence of the freshwater may lead to similar sanitation problems as those that caused the deaths of 3.4 million people due to waterborne diseases in 2005 [2]. Important causes of water shortages include the acceleration of industrialization, increasing water demands, and severe climate changes as a result of global warming.

In order to overcome, or at least mitigate, these water scarcity problems, seawater can and should be considered as an alternative freshwater source. Seawater comprises the majority of the water present in the world, though it is not suitable for direct human consumption and/or industrial and agricultural use. To this end, desalination, which is the process of removing of salts and other minerals from feedwater, can offer a solution for achieving the required quality and quantity of freshwater. Desalination technologies can produce large amounts of water over a given time, and the construction period of desalination plants is relatively short compared to other technologies; therefore, a considerable amount of water can be produced relatively quickly. Moreover, some desalination technologies are independent of seasonal variations and weather conditions. Thus, the desalination industry has been rapidly developing over the past couple of decades, with the desalination market showing a corresponding growth (Fig. 1) [3].

Desalination can be divided into two primary technologies: thermal and membrane processes. Thermal methods are based on an engineering approach to replicating the earth's natural water cycle. The basic principle of the thermal method is to boil raw water and then condense it in order to get product water (distillate); MSF and MED are the two most dominant types of thermal desalination processes. Thermal methods can produce a high quality of product water and deal with the demands of large-scale production; however, these methods consume a large amount of energy. Typical membrane desalination processes using thermal methods are based on the characteristics of semipermeable membranes, with RO being the most popular. Since membrane desalination technologies do not utilize fossil fuels, like thermal processes do, they are referred to as energy-efficient desalination processes [4].

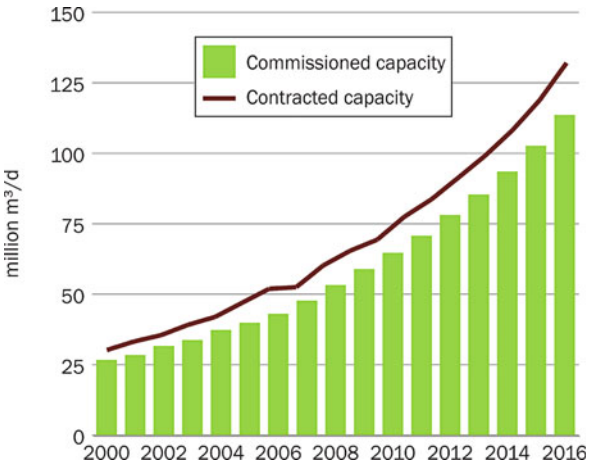
In order to meet their freshwater demands, a number of countries already rely on desalination technologies, especially in the Middle East and some parts of Europe. Figure 2 summarizes the capacity of seawater desalination plants in the world, based on typed over a period from 1981 to 2011 [5]. Until 2008, thermal methods dominated the market, mainly due to their use in fossil fuel–rich Middle Eastern countries. Other regions started to construct large-scale desalination plants about 10 years ago, and since that time, the method for desalination shifted from thermal to membrane-based processes. Indeed, the market share of membrane desalination processes has been increasing in the world, parallel with the water demand (Fig. 2) [3, 4].

This text includes chapters on thermal desalination (MSF and MED), membrane desalination (RO, FO, and MD), and future directions. In addition, each chapter consists mainly of sections on (1) process principles, (2) system investigation, (3) system model description, and (4) the advantages and limitations of each process.

Multistage Flash (MSF) Distillation

Principle

The basic principle of the multistage flash (MSF) distillation process is flash evaporation, which separates volatile components from a solution. When the operating pressure rapidly decreases below the saturated



Desalination Technology for Sustainable Water Resource. Figure 1
Desalination market forecast; contracted and commissioned capacities [3]

vapor pressure at a given temperature, the latent heat of vaporization is removed from the contained heat of the liquid and flash evaporation occurs. In this process, the temperature of the inlet solution should be higher than the boiling point of the solution in the evaporation space; hence, a portion of the inlet solution can be vaporized based on the adiabatic equilibrium. Therefore, complete evaporation in the MSF process can be achieved by simultaneously increasing the temperature and decreasing the pressure.

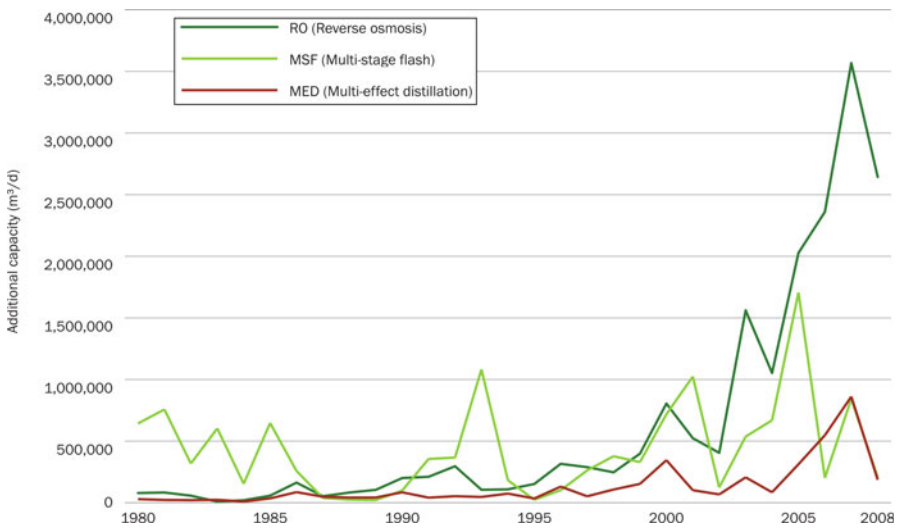
System Investigation

MSF desalination plants consist of several flash chambers in series. The basic principles of the MSF desalination plant can be classified into three parts: heat input, heat recovery, and heat rejection (Fig. 3).

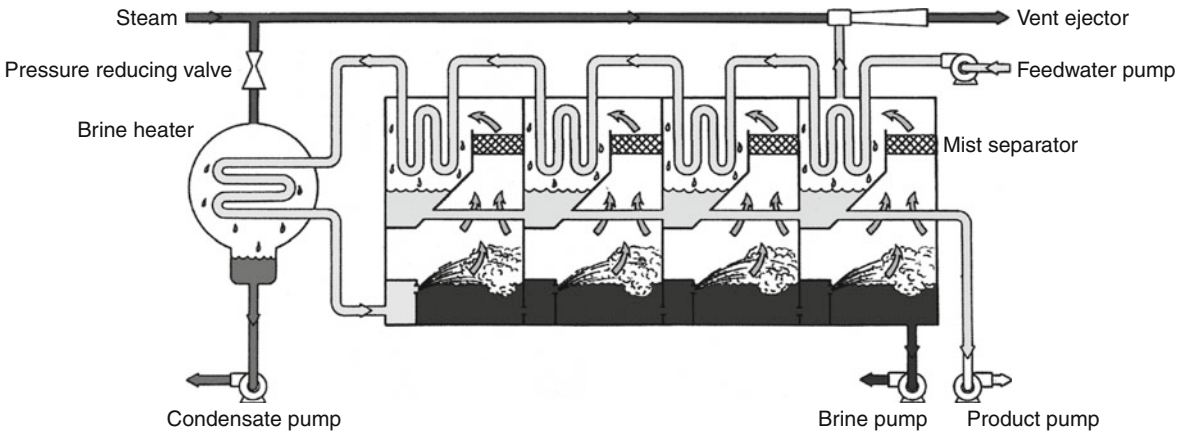
In brief, seawater is heated for the heat input, which is performed in a brine heater under low pressure conditions. The operating temperature of the boiled brine is generally kept between 90°C and 120°C in order to avoid scale formation. However, even though the system efficiency can be increased by using higher temperatures, this also increases the potential of scale formation. Scaling can accelerate the corrosion rate of devices, especially for parts that are in direct contact with the seawater. Note that a large-scale MSF desalination plant typically consists of 19–28 stages.

Seawater passes through the heat exchanger in the upper side of the flash chamber and flows into the brine heater. The boiled brine is then jetted through the orifice into the lower pressure chamber and vaporized. Any surplus energy can then be transferred to the partial latent heat; therefore, no additional external heat is required. The heating process continues until the seawater temperature reaches the boiling point of the flash chamber. In other words, flash distillation occurs based on the sequential reduction of pressure on the heated seawater. Further flash evaporation occurs when the liquid brine passes from one stage to the next lower pressure stage.

In order to minimize the loss of droplets from the brine, each evaporator stage is equipped with demisters. In this type of system, the evaporator is typically installed in conjunction with a decarbonator and deaerator in order to remove carbonate and eliminate dissolved gases. The decarbonator converts bicarbonate in the seawater to carbon dioxide by adding acid to the solution; bicarbonate is the main species in the seawater that leads to scaling. A steam jet ejector system then maintains the evaporation stage by extracting noncondensable gases from the system. The vaporized water is subsequently condensed using heat exchangers in order to produce distilled water. In these heat exchangers, latent heat from the flashed vapor is used to heat the inlet brine. The water quality produced from an MSF process normally contains around 2–10 ppm dissolved solids; as such, remineralization might be



Desalination Technology for Sustainable Water Resource. Figure 2
Desalination process trend by contracted capacity [5]



Desalination Technology for Sustainable Water Resource. Figure 3
Schematic of multistage flash (MSF) distillation [3]

required for potable water usage as the distilled water contains very low concentrations of salts [3, 4, 6].

System Model

Changes in the density of the cooling brine are negligible; therefore, the mass holdup of the cooling brine can be considered constant. In addition, as the heat transfer coefficient will be much larger for the condensation of vapor outside the tubes than inside the tubes, the tube temperature is assumed to be equal to the vapor temperature. Based on the above assumptions, the model equations are as follows [7]:

Mass and energy balance for the flash chamber

$$(M_b/1000)\frac{dC_{b,out}}{dt} = B_{in}(C_{b,in} - C_{b,out}) + VC_{b,o} \tag{1}$$

$$(M_b/1000)\frac{dh_{b,out}}{dt} = B_{in}(h_{b,in} - h_{b,out}) + V(H - h_{b,out}) \tag{2}$$

$$M_b = \rho_b A_b \cdot LE \tag{3}$$

where M, C, B, and V are the mass (kg), salt concentration, flashing brine flow rate (t/h), and vapor flow rate (t/h), respectively. In addition, h, H, and ρ are the liquid specific enthalpy (kcal/kg), vapor specific enthalpy (kcal/kg), and density (kg/m^3); A is the heat transfer area (m^2), and $LE$is the brine level (m). The subscripts in, out, and b indicate the input, output, and flashing brine.

Based on the assumption that the contents of the flash chamber are well mixed, Eqs. 1 and 2 can be rewritten. Therefore, the temperature, concentration, and specific enthalpy of the flash chamber will be equal to those of the outlet brine from the chamber.

Mass and energy balance for the brine heater

$$Q = 1000S(H_S - h_S) \tag{4}$$

$$M_{T,H}C_{P,ave}\frac{dT_F}{dt} = Q + 1000(F_{in}h_{F,in} - F_{out}h_{F,out})$$

$$F_i = F_o \tag{5}$$

$$M_{T,H} = \rho_F\left(\frac{\pi ID^2}{4}\right)_H L_H \cdot NT_H \tag{6}$$

where Q, S, and T are the heat transfer rate (kcal/h), steam flow rate (t/h), and temperature (°C), respectively. Also, C_p is the specific heat (kcal/kg°C) and F is the cooling brine flow rate (t/h); ID, L, and NT refer to the inner tube diameter (m), tube length (m), and number of tubes. The subscripts T, H, and F refer to the tubes, brine heater, and cooling brine.

Mass and energy balance for the condenser

$$D_{out} = D_{in} + V \tag{7}$$

$$T_D = T_b + T_L \tag{8}$$

where T_L is the total temperature loss caused the boiling point elevation, nonequilibrium, and pressure loss in the demister. Then,

$$D_{out}h_{D,out} = D_{in}h_{D,in} + Vh_D \tag{9}$$

where D is the distillate flow rate (t/h) and the subscript D is the distillate [7].

Advantages and Limitations

Since the core technologies of MSF desalination are well established, the application of MSF desalination is relatively easier than for other technologies. However, there is still a need for further development of materials and process design. In addition, by improving the process efficiency, the MSF life cycle can be extended by up to 30 years. Over the last several decades, MSF has been the dominant desalination technology; therefore, there is a corresponding strong background understanding of construction, maintenance, and operational expertise.

In spite of these advantages, however, MSF desalination has several limitations that need to be addressed. Recent increases in oil prices are an important limitation as they may lead to an inability to pay for MSF operations. In addition, the MSF desalination market is occupied by few major engineering companies, which limits the choice of alternative technologies to be introduced [3, 4] (Table 1).

Multieffect Distillation (MED)

Multieffect distillation (MED) is one of the oldest desalination processes, developed in the early nineteenth century. The MED process takes place in a series of evaporators, which are called effects. Similar to MSF, MED also uses the principle of the ambient pressure reduction; however, even though the MSF process replaced MED after its development, there has been a resurgence MED usage as new design approaches have been developed [3, 4, 8].

Desalination Technology for Sustainable Water Resource. Table 1 Advantages and limitations of MSF desalination process

Advantages	Limitations
Mature technology	Heat source cost
Longevity	Electrical consumption
Process simplicity for operating	Requirement to condensate
High product water quality	Higher energy costs
Easy to scale up (large scale)	Duopoly of the market
Requires minimum pretreatment	

Principle

The MED system consists of multiple evaporators in a sequence, which operates at progressively lower pressures. The vapor is generated from seawater within each effect, which is then transferred to the next effect as a heat source. While this vapor is being used as a heat source, product water condenses within the effect. A key benefit to this process is that once the vapor is generated at the first stage, there is no additional heat requirement for the rest of the process. The pressure and temperature sequentially decrease for each effect, from the first to the last stages [8].

System Investigation

After seawater enters the first effect, its temperature increases up to the boiling point. Cool seawater is then sprayed over a steam pipe to initiate the condensation process of the steam inside the pipe. At the same time, thin seawater layers boil outside the pipes as they absorb heat from the steam pipe. Freshly produced steam is then introduced into the pipes in the next effect, and this is repeated throughout the process. More recently, the introduction of vapor compression, which energizes low pressure steam, has been found to increase the overall efficiency of the MED process [3] (Fig. 4).

System Model

A "rigorous" model of MED columns includes the energy, material (overall), and component material balances. The balance equations for the distillation column are written as follows [9]:

Overall material balance

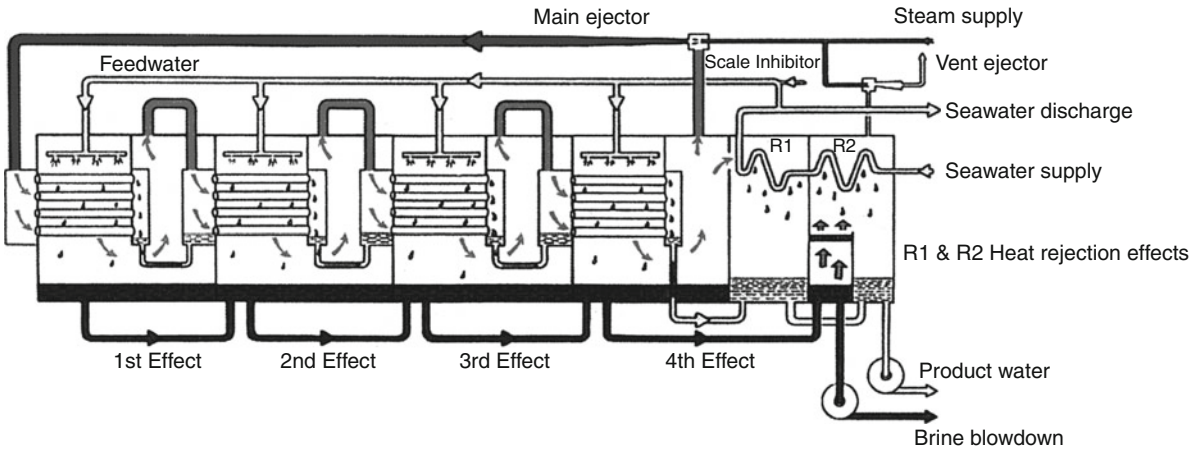
$$\frac{d\left(M_{\mathrm{L},i}+M_{\mathrm{V},i}\right)}{dt}=L_{i+1}+V_{i-1}-L_i-V_i \tag{10}$$

Component material balance

$$\frac{d\left(M_{\mathrm{L},i}x_{ij}+M_{\mathrm{V},i}y_{ij}\right)}{dt}=L_{i+1}x_{i+1,j}+V_{i-1}y_{i-1,j}-L_ix_{i,j}-V_iy_{i,j} \tag{11}$$

Energy balance

$$\frac{d\left(M_{\mathrm{L},i}u_{Li}+M_{\mathrm{V},i}u_{V,i}\right)}{dt}=L_{i+1}h_{\mathrm{L},i+1}+V_{i-1}h_{\mathrm{V},i-1}-L_ih_{\mathrm{L},i}-V_ih_{\mathrm{V},i} \tag{12}$$



Desalination Technology for Sustainable Water Resource. Figure 4
Schematic of the MED process [3]

Here, $M_{L,i}$ and $M_{v,i}$ are the holdups in the liquid and vapor phases on stage *i*. In addition, index *j* denotes the component *j*, L and V are the liquid and vapor flows, x and y are component fractions in the liquid and vapor phases, and h_L and h_V are the liquid and vapor enthalpies.

Advantages and Limitations

Although MED is the oldest desalination technology, new developments for process efficiency, such as new design approaches, have rejuvenated the interest in MED applications. Even though MED technology is more complex than MSF – even with current developments – MED remains competitive with MSF. However, one of the major limitations for MED plant installation is the scarcity of titanium, which is used as a construction material [3, 4].

Desalination Technology for Sustainable Water Resource. Table 2 Advantages and limitations of MED desalination process

Advantages	Limitations
Rejuvenated technology	Complex technology
Competitive on a larger scale	Small unit size
High product water quality	Non-cost-competitive with RO process
Less cooling water required than MSF	Lack of track record
Lower capital cost than MSF	Lack of competition
Well established in certain markets	Titanium scarcity

Reverse Osmosis (RO)

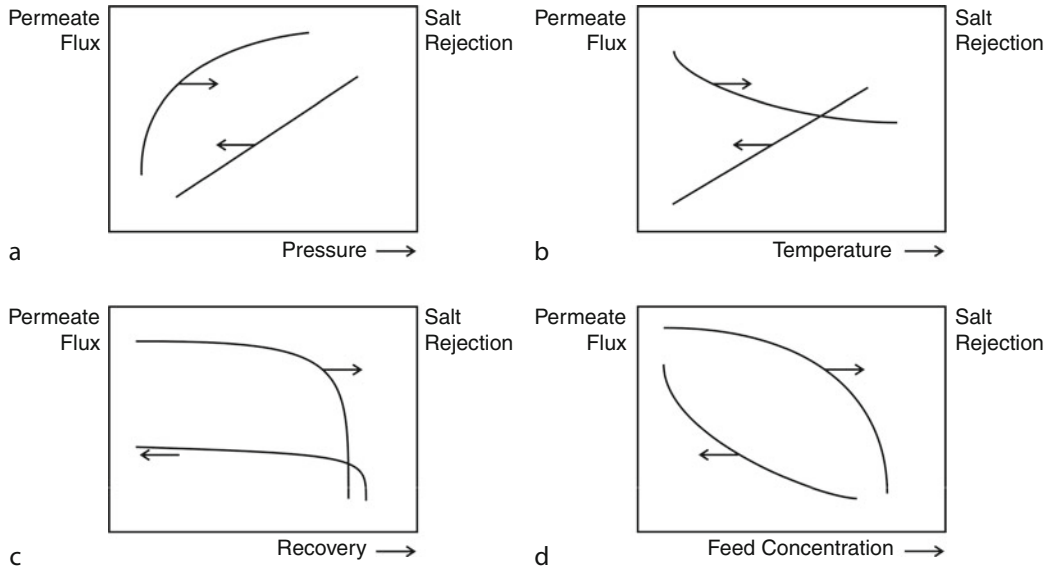
Principle

Reverse osmosis (RO) is an unnatural process, which reverses water flow in an osmotic system. RO requires a larger transmembrane pressure than the typical transmembrane osmotic pressure used to push water from a high concentration zone to a low concentration zone through a semipermeable membrane [10]. The semipermeable membrane allows a solvent to pass to the permeate side, whereas solutes are rejected by the membrane. Note that RO processes can be categorized into two types according to their pressure strength: (1) high-pressure RO, which requires a pressure between 5.6 and 10.5 MPa, and can be used for seawater desalination; and (2) low-pressure RO, which requires a pressure between 1.4 and 4.2 MPa, and can be used for brackish water desalination. Both RO processes provide a high rejection efficiency (95–99%) of sodium chloride (NaCl) (Table 2).

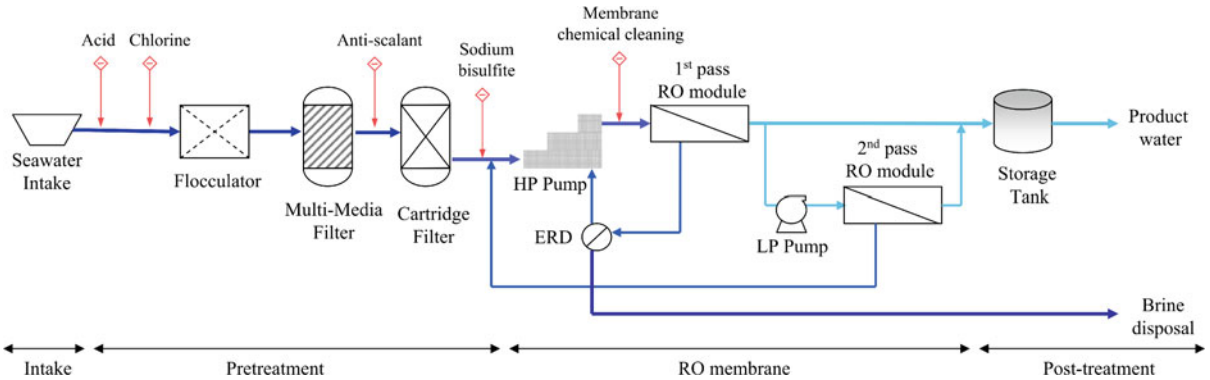
Generally, RO membrane configurations can also be categorized into two types: flat (e.g., plate-and-frame and spiral-wound modules) and tubular (e.g., tubular, capillary, and hollow-fiber modules) [11]. However, the spiral-wound type has been more widely selected, due to its lower maintenance cost and simpler

design than is the case for other configurations [12]. The spiral-wound type consists of a central tube (used to collect the permeate water), membrane sheets (which are rolled onto the central tube), and a pressure vessel (which covers the system to ensure a stable pressurized environment) [13]. In practical applications, several spiral-wound elements are connected in series in one pressure vessel in order to increase the RO process efficiency.

To evaluate the process performance, recovery and rejection rates are broadly used as key parameters [14]. The performance of an RO membrane process is principally governed by site-specific variables (e.g., solute concentration, temperature, and pH), membrane variables (e.g., membrane type, module geometry, and module arrangement), and process variables (e.g., feed flow rate, operating pressure, and operating time) [15]. Moreover, the design, operation, and maintenance of an RO plant also influence the performance of the RO process. In order to maintain the feed flow rate, a high-pressure pump and a flow-regulating valve on the brine side can be used. The impact of each parameter on the membrane performance is shown in Fig. 5 [12]. Figure 6 then presents the schematic of a typical two-pass



Desalination Technology for Sustainable Water Resource. Figure 5
Impact of several factors affecting membrane performance [12]



Desalination Technology for Sustainable Water Resource. Figure 6
Schematic of a typical SWRO desalination process [16]

seawater reverse osmosis (SWRO) desalination process, which includes a seawater intake, a pretreatment system, an RO system, and a posttreatment system.

Instead of using high-pressure RO systems for seawater desalination, nanofiltration (NF) can be considered as an alternative process as it has a lower pressure requirement. NF is also a pressure-driven membrane process and has a pore size between RO and ultrafiltration (UF) [17]. NF membranes have a molecular weight cutoff (MWCO) of about 200–1,000 Da, which make NF processes suitable for separating dissolved components with a molecular size of about 1 nm [18]. Separation mechanisms in the NF process are based on the sieving effect, solution-diffusion, Gibbs–Donnan effect, dielectric exclusion, and electromigration. Therefore, it is possible to separate both charged and uncharged organic solutes using NF [7]. NF also demonstrates a high rejection for divalent/multivalent ions, as well as for organic compounds that have molecular weights above 300 g/mol [17]. However, NF shows a low rejection efficiency for monovalent ions and nonionized organics, especially for those with a low molecular weight. Nonetheless, according to previous studies, NF membrane processes can be successfully applied to seawater desalination. For instance, results have shown that NF desalination could produce a water product that contains less than 1% salt content [19].

System Investigation

An SWRO distillation system generally consists of: an intake structure, a pretreatment process, an RO filtration process, a posttreatment process, and optionally an energy recovery process. System investigation for SWRO is important because it is helpful not only to increase the system efficiency but also to reduce the unit production costs. Moreover, the construction of large-scale RO plants can also reduce the cost per production unit. Thus, several large-scale plants have frequently been reported as effective and productive case studies. Table 3 presents a summary of the ten largest SWRO plants [20], where it should be noted that there are several issues that need to be evaluated prior to setting up a large-scale SWRO process.

Intake Structure and Site Area In order to ensure a high level of seawater quality, a subsurface seawater intake at 10–15 m below the sea surface is required [14]; therefore, the cost for logistics and construction will increase. Based on this requirement, desalination plants that are located close to the intake site might be considered to be more cost efficient.

Desalination Technology for Sustainable Water Resource. Table 3 Ten largest SWRO plants in the world [20]

Country	Location	Capacity (m^3/h)	Year of construction	Membrane manufacturer	Module
United Arab Emirates	Fujairah	7,083	2004	Hydranautics/Nitto	Spiral wound
Saudi Arabia	Median/Yanbu	5,333	1998	Toyobo	Hollowfilber
Spain	Carboneras	5,000	2003	Hydranautics/Nitto	Spiral wound
Trinidad and Tobago	Point Lisas	4,542	2002	Hydranautics/Nitto	Spiral wound
USA	Tampa Bay	3,917	2003	Hydranautics/Nitto	Spiral wound
Saudi Arabia	Al Jubail	3,750	2002	DuPont/Toray	Hollowfilber/spiral wound
Spain	Cartagena	2,708	2002	Hydranautics/Nitto	Wickelelement
Saudi Arabia	Jeddah I	2,367	1989	Toyobo	Hollowfilber
Saudi Arabia	Jeddah II	2,367	1994	Toyobo	Hollowfilber
Spain	Marbella	2,350	1998	DuPont	Hollowfilber

Raw Seawater Conditions Since raw seawater conditions depend on the specific location, it is important to investigate individual parameters when selecting the pretreatment process, RO system design, membrane material, and chemical cleaning methods [21]. Thus, it is necessary to continuously monitor the raw seawater conditions, in order to make prerequisite adjustments pertaining to the pretreatment process. A summary of raw seawater compositions from several sites in the world are given in Table 4 [22–31].

Pretreatment Processes

- *Conventional approach*
  As mentioned above, the seawater quality is a very important factor in RO process design and operation; since the plant operator cannot control the seawater quality, an effective pretreatment process is critical for improving overall SWRO system performance [27]. The basic components for the pretreatment process include screening, coagulation, flocculation, media filtration, and a cartridge filter. Table 5 summarizes several key issues that need to be considered during the pretreatment process [14, 25, 32, 33].
- *UF membrane approach*
  Ultrafiltration (UF) pretreatment processes are an alternative pretreatment method for the further improvement of feedwater quality. The major disadvantage of conventional pretreatments is the utilization of chemical agents, which tend to shorten the membrane lifetime [14, 26, 34, 35]. Therefore, UF can be a good candidate for SWRO processes. Table 6 presents a summary of advantages of UF over conventional pretreatments.

  Table 7 can also be used as a decision tree for the pretreatment selection for SWRO desalination systems.

RO Filtration Processes The major factors that can significantly affect the cost and performance of an RO membrane process are described as follows [36–48].

- *Permeate recovery and salt passage*
  Permeate recovery and salt passage rates are significant parameters that are commonly used to evaluate RO membrane performance. A high-efficiency RO membrane process should provide a high permeate recovery and low salt passage rates, which can be controlled by improving the membrane properties.
- *Concentration polarization*
  Concentration polarization affects solute adsorption and gel layer formation onto the active sides of the membranes. These phenomena increase the osmotic pressure and decrease the permeate flux. Polarization layers are formed by hydrophilic macromolecules, whereas gel layers are formed by hydrophobic macromolecules; both layers could lead to a severe flux decline.
- *Membrane fouling*
  Membrane fouling is an important factor that has an adverse effect on RO process performance. Membrane fouling occurs via the deposition of particles onto the membrane surface or into membrane pores. This phenomenon leads to an increase in the transmembrane pressure (TMP), followed by a decrease in the quantity and quality of the water product [49–51]. And although pretreatment processes are intended to remove majority of the foulants from raw seawater, pretreated water can still contain foulants such as dissolved organic compounds and tiny colloidal particles [52]. To date, numerous research groups have investigated the fouling mechanisms in attempts to predict fouling formation [53–56]; however, they could not provide a full explanation and/or solution for fouling occurrences. Therefore, membrane fouling remains as a significant factor limiting the further development of RO membrane processes. To this end, the major foulants are natural organic matter (NOM), colloidal particles, microorganisms, and inorganic particles. Scaling is the major inorganic fouling component, which forms by the precipitation of complex salts such as calcium carbonate ($CaCO_3$), calcium sulfate ($CaSO_4$), silica (SiO_2), and iron hydroxide ($Fe(OH)_3$), some of which occur during chemical pretreatment processes. Therefore, minimizing the chemicals used during pretreatment is required in order to reduce the fouling rate.
- *Chemical cleaning*
  Chemical cleaning is an inevitable process for RO in order to eliminate fouling and/or scaling on the membrane. Alkaline (NaOH), acid (citric acid or H_2PO_4), ethylenediaminetetraacetate (EDTA), chlorine (Cl_2), and surfactants/detergents are

Desalination Technology for Sustainable Water Resource. Table 4 Compositions of raw seawater at several sites in the world

	pH	TDS (mg/l)	Temp (°C)	Total alkalinity (mg/l)	Bicarbonate HCO_3^- (mg/l)	Sodium Na^+ (mg/l)	Potassium K^+ (mg/l)	Magnesium Mg^{2+} (mg/l)	Calcium Ca^{2+} (mg/l)	Fluoride F^+ (mg/l)	Chloride Cl^- (mg/l)	Sulfate SO_4^{2-}(mg/l)	Silica SiO_2^-(mg/l)	Barium Ba^{2+}(mg/l)	Phosphate PO_4^{3-}(mg/l)
Kifan site 1 [22]	7.21	11,435	28.6	133	133	2,462	39	303	693	2.15	4,013	3,016	2	0.18	<0.05
Kifan site 2 [23]	7.15	10,786.6	–	143.3	143.3	3,045	55.2	432.4	869.9	1.95	3,983.8	3,018.9	5.45	<0.05	<0.05
Canary Islands [30]	–	–	–	6,600	–	9,055.5	251.8	1,354.3	205.5	–	21,328.5	3,300	–	–	–
Crude seawater (Canary Islands) [25]	–	–	–	–	171.9	11,810	–	1,438.7	448.9	0.5	21,317.2	2,900	–	–	–
Atlantic Ocean [31]	8	40,077	17–20	162		12,255	441	1,477	464	1.5	22,019	3,075	1.4	0.4	–
Jeddah, Saudi Arabia (Red Sea) [28]	8	41,200	32	–	–	–	–	1,464	520	–	22,000	–	0.4	–	–
Yanbu City [26]	8.1–8.3	41,300–46,400	25	120–130	–	11,700–12,500	425–650	1,500–1,600	490–560	1.5	21,600–23,500	3,000–3,200	0.5	–	<0.1
Tajura, Libya [24]	8.3	38,000	–	–	163	11,600	419	1,427	455	–	20,987	2,915	–	2	–
Bocabarranco [29]	–	38,000	–	–	250	11,415	450	1,520	450	–	20,800	3,110	5	–	–
Surface seawater (open intake) [27]	8.07	–	19.3	–	161	10,945	410	1,471	441	1.55	20,900	2,965	0.2	10	0.02
Well seawater [27]	7.8	–	21.1	–	166	10,747	399	1,340	438	1.35	20,100	2,840	1.5	20	0.09
Combined water [27]	7.87–7.92	–	18.6–20.5	–	162	10,841	402	1,355	441	1.35	20,700	2,855	0.9	15	0.05

commonly used to eliminate NOM, inorganic scalants, scalants (e.g., $CaSO_4$), biofouling (biofilm), and colloids, respectively. In order to maximize the pretreatment efficiency, the type and concentration of the cleaning agent need to be considered. As an example, membranes can be cleaned via the cleaning-in-place (CIP) method, though this method requires information pertaining to interactions between the foulants/scalants and the membrane surface [39]. In CIP, the cleaning priority for essential foulants is as follows: silica colloids > adsorbed organic compounds > particulate matter (iron and aluminum colloids) > microorganisms > metallic oxides.

Energy Recovery Processes Since it is possible to transfer energy from the high-pressure brine stream to the feed stream [57], a suitable energy recovery process can be used. Current energy recovery technologies can allow us to reduce the total energy cost by up to 40% [20]. Typically, there are three types of energy recovery devices in SWRO systems [58–61]:

1. A Pelton wheel turbine (PWT) with an efficiency rate range of 40–60%
2. A pressure exchanger (PX) with a maximum efficiency rate as high as 95%
3. A hydraulic turbo charger (HTC) with an efficiency rate ranging from 50% to 65%

According to existing data, the PX shows the highest efficiency and most dynamic stability [59].

Posttreatment Processes In SWRO desalination processes, posttreatment is generally necessary in order to improve the quality of product water to meet potable water quality standards. Posttreatment processes include pH adjustment, minimal remineralization, disinfection, and boron removal [20, 62, 63]. The pH of desalinated water should also be adjusted to a range of 6.8–8.1, and disinfection techniques need to be applied to remove bacteria or other organisms from the product water. However, the concentration of disinfection by-products (DBPs) and bromate (BrO_3^-) in the final product should not exceed potable water quality standards; the concentration of boron should also meet these same standards. Previous studies reported that SWRO membrane systems could reject boron in a range from 92% to 94%; however, the boron rejection efficiency is strongly dependent on the pH level of the water. It was further reported that at a pH less than 9.5, approximately 50% of boron

Desalination Technology for Sustainable Water Resource. Table 5 Key issues to be considered in the pretreatment process

Foulant/parameter	Chemical
Microorganism	Sodium hypochlorite (NaOCl), Cl_2, $KMnO_4$, or O_3 can be used to control biofouling
	H_2SO_4 can be applied to assist the biocide action of NaOCl
Inorganic particles	Sodium hexametaphosphate $(NaPO_3)_6$ is usually added in doses to control scaling
	H_2SO_4 can be used to assist the action of scale inhibitors
pH adjustment	H_2SO_4 can be added to regulate the pH for polyamide-type RO membranes
Colloidal particles and dissolved organics	Ferric or alum salts are often used to coagulate and flocculate colloidal particles and dissolved organics
Suspended solids	Ferric or alum salts are often used to coagulate and flocculate colloidal particles and dissolved organics
	Anthracite (~1 mm) is often applied during the granular media filtration process to remove suspended solids
	A media backwashing process with air is followed by the granular media filtration process to remove particles captured in the filters
Particulate matter	Cartridge filters are usually tasked with preventing the sudden appearance of particulate matter
Neutralizing the residual active chlorine	Sodium metabisulfite ($NaHSO_3$) is primarily used to neutralize residual active chlorine, especially for polyamide-type RO membranes

Desalination Technology for Sustainable Water Resource. Table 6 Comparison of conventional and UF membrane pretreatments

	Conventional pretreatment	UF membrane pretreatment
Treated water quality	Unstable and fluctuating water quality depending on raw seawater (silt density index, SDI < 4.0)	Stable and constant water quality (SDI < 2.0)
Average RO flux	100%	20% higher
RO membrane fouling rate	High fouling potential	Lower fouling potential
RO membrane cleaning frequency	1–2 times per year	4–12 times per year
Typical lifetime	Filters: 20–30 years	UF/NF membranes: 5–10 years
	Cartridges: 2–8 weeks	Cartridges: not often needed
RO membrane replacement rate	100%	33% lower
Capital cost	100%	0–25% higher
Footprint	100%	30–60% smaller
Energy consumption	Lower than UF	Higher than conventional
Chemical dosing rate	High	Lower
Intake line	Long	Shorter
Operation/management cost	High	Lower
Miscellaneous	–	Better boron control

Desalination Technology for Sustainable Water Resource. Table 7 SWRO pretreatment process selection

Water quality process design (selection)	SDI < 4	SDI > 4				
	NTU < 0.5	$0.5 \leq$ NTU < 2	$2 \leq$ NTU < 20	$20 \leq$ NTU < 40	$40 \leq$ NTU < 100	$100 \leq$ NTU
Coagulation and flocculation		X	X	X	X	X
Sedimentation or DAF				X	X	
Enhanced sedimentation or DAF						X
Single-stage granular media filtration	X[a]	X[a]				
Two-stage granular media filtration			X[a]	X[a]	X[a]	X[a]
MF/UF membrane filtration	X[a]	X[a]	X[a]	X[a]	X[a]	X[a]

[a]Select either granular media or membrane filtration. For turbidity >20 NTU, consider using a combination of both

could be removed, and that at 10.5, the removal rates increased up to 100%. In addition, membrane material can also play an important role in the boron rejection rate. For example, polyamide-type membranes showed higher rejection rates for pH values of less than 9.5, though the most effective posttreatment for boron removal was found to be to use boron selective resins (BSR) because they are not affected by operating conditions such as temperature, pH, and/or salinity [64–66].

The Marin Municipal Water District (MMWD) SWRO pilot system in the city of Corte Madera in Sonoma County, California, developed the following general water quality objectives for after posttreatment stabilization (Table 8) [63].

Brine Disposal and Environmental Impact Brine from SWRO processes contains a high concentration of salts (TSD ~70,000 mg/L), in addition to remnants of chemical product/by-products, which are used in overall SWRO process. The effect of discharging this high concentration of salts and other chemicals into the environment should be seriously considered as these high concentrations could inhibit a natural mixing caused by density differences. Moreover, a leak of the brine pipeline into underlying aquifers could release heavy metals and cleaning agents into the environment, potentially leading to serious groundwater contamination. As such, environmentally friendly brine disposal methods should be selected, including methods with low carbon dioxide production and low energy consumption. Several options of brine disposal include: deep well injection into nonpotable aquifers, utilization of evaporation ponds (zero discharge), and connecting the brine pipeline to wastewater treatment plants and then treating it together with other discharged water [20, 30, 67–69].

System Model

Recently, desalination technology has shown remarkable improvements in many aspects: membrane material, equipment, and cost and energy reductions [70, 71]. However, in order to reduce the overall process cost, optimization of process operation and maintenance are required, as well as advanced control of the system. Therefore, a modeling approach is required for the system analysis of SWRO and also for the optimization of the technologies, in order to achieve an effective cost reduction.

A strong modeling approach can be used to predict performance and optimize process operations and process controls. Usually, advanced statistical models (e.g., artificial neural networks and genetic programming) are broadly applied because of their simplicity and user-friendly interfaces. And since these models are based on the input/output data, they do not require huge data sets. They have also been proven useful when a physical understanding is inadequate. For example, deterministic models, which are developed based on physical laws, can give a better physical insight of the system process. These models have also been used to provide a real time insight into process behavior monitoring, such as for state and measured variables.

Desalination Technology for Sustainable Water Resource. Table 8 Proposed general water quality objectives for after posttreatment stabilization for the MMWD SWRO pilot system in the city of Corte Madera in Sonoma County, California [63]

Parameter	Units	MMWD-treated reservoir			Sonoma County water			SWRO pilot objectives		
		Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
TDS	mg/L	119	136	86	171	186	148	120	180	60
Hardness	mg/L	62	74	52	105	112	96	60	110	60
Alkalinity	mg/L	61	70	49	119	125	110	60	110	50
pH	Units	7.8	7.9	7.8	8.1	8.4	7.8	7.9	8.2	7.8
Color	CU	<3	<3	<3	<3	<3	<3	<3	<3	–
TOC	mg/L	1.6	2.4	1.1	0.9	1.2	0.7	<1	1	–
Sodium	mg/L	16	25	11	20	23	16	30	50	10
Chloride	mg/L	27	37	22	8	10	7	50	70	10
Boron	mg/L	<0.05	<0.05	<0.05	0.28	0.26	0.16	0.3	0.5	–
SAR	–	–	–	–	–	–	–	3	6	–

Pretreatment Process Models There have been several models developed for UF/MF pretreatment and posttreatment processes. These existing models are mostly described in a similar manner, based on mathematical approaches that can be applied to the specific desired process.

- *UF/MF*
  Microfiltration and ultrafiltration both resemble conventional coarse filtration; therefore, they are suitable for retaining suspensions and emissions. Darcy's law can be used to describe volume flow in UF/MF processes. Furthermore, for flux estimations, both the Hagen–Poiseuille and the Kozeny–Carman equations can be applied when laminar convective flow assumptions are valid [72].
- *Granular media filter*
  Models for granular media filters require a full understanding of particle transport and particle deposition phenomena in order to predict the fate of the colloidal particles. In this model, the main transport mechanisms of colloidal particles, from pore fluid to the surface of filter grains, are gravitational sedimentation and Brownian diffusion [73]. The first filtration model that was applied to a water treatment system was introduced by Yao et al. [74], though this model assumed that viscosity and van der Waals interactions have no effect on the system. Therefore, a prediction model for colloidal filtration must be developed based on a numerical approach that includes a convective-diffusion equation [75].

  However, a comprehensive solution of the convective-diffusion equation is not yet readily available. Thus, a semiempirical correlation model for the single-collector contact efficiency in granular filtration was introduced by Rajagopalan and Tien [76]. This model was developed by considering the effects of viscous and van der Waals interactions [73], which then was able to provide more realistic predictions than the original model. Recently, this model has been broadly applied to the prediction of single-collector contact efficiency in a granular filtration system [77].

SWRO Filtration Process Models There are three types of deterministic models for RO filtration processes: an irreversible thermodynamics-based model, a diffusion-based model, and a porous model [15]. These models have all been applied in attempts to evaluate the solvent and solute transport in RO filtration processes based on the membrane characteristics. RO transport models can also be classified into three categories: solution-diffusion (SD) models, irreversible thermodynamic (IT) phenomenological models, and pore-based models [15]. It should be noted, however, that numerous research groups have studied the solute transport mechanisms in RO processes more than the solvent transport because the solute flux is not proportional to the net driving pressure (NDP) [78]; several groups have also investigated membrane fouling via membrane resistance-in-series-based simulations [37, 55, 79, 80].

- *SWRO operational models*
  Operational models for SWRO processes can help to increase overall process efficiencies and production rates, and be used to determine the economic feasibility of the process. In additional, these models can also be used to improve the process management.

  The cost and total annual profit (TAP) estimations of an SWRO process are two of the most important criteria for establishing a feasible desalination process. Developed package models must be proven satisfactory in terms of permeate water flow rate and concentration, which can then be extended to estimate the operational cost of the system. Moreover, the results of cost estimation show that the models can successfully predict the TAP, according to operation time, feedwater TDS, and permeate water TDS [81].

  Most SWRO plants operate in isobaric conditions, which significantly reduce their ability to respond to an external disturbance. However, the permeate salinity significantly depends on the feed salinity and a change in the feed salinity can affect both process efficiency and operational costs. Therefore, a cost-effective SWRO process can be utilized by using a nonisobaric pressure control that considers possible fluctuations in the feed seawater concentrations [82]. Another study, which also investigated nonisobaric pressure control together with feed temperature control, showed that feed pressure required can be reduced by approximately 10 bars in an SWRO–MSF hybrid system [83].

An artificial neural network (ANN) model has been used to optimize the cost effectiveness of an SWRO–MSF hybrid system [84]. This study showed that feedwater temperature manipulation can directly improve the system performance, based on a linear increase with respect to the feed temperature over a 1-year period. Another SWRO–MSF hybrid system study also used feedwater temperature manipulation to increase the permeate flow rate [85]. It was seen that feedwater temperature manipulation can cause an increase in the permeate flow rate by up to 0.2% per day.

Lee et al. [86] proposed a membrane resistance model that utilizes temperature and pressure correction factors in order to evaluate membrane performances. The developed model requires a minimum parameter input to investigate the performance of the membranes under different conditions. This model can be used to suggest appropriate operating conditions for the desired water production.

- *SWRO network model*

Representations of RO network (RON) models developed by El-Halwagi [87] are now widely used to optimize the RO plants under various conditions [88–91]. In addition, RON models that can obtain realistic and economical solutions by optimizing total costs [41, 92] have also been developed.

A chronology of system engineering approaches for the modeling and optimization of the RO systems is given in Table 9.

The structure of an extended system analysis may start with the design of the seawater quality and the membrane modules. Then, a cost-effective RON configuration (RO modules, pumps, and energy recovery devices), the operating condition, and the optimal arrangement of membrane elements can be analyzed. The structure of this type of RO system analysis was first suggested and used by El-Halwagi [87] and Voros [90, 93]. Figure 7 shows a simplified RON representation that Lu et al. [94, 95] have recently developed by applying a stream split ratio, isobaric-mixing constraints, and a PX energy recovery device to the previous RON model.

Desalination Technology for Sustainable Water Resource. Table 9 Chronological systems engineering approaches for RO membranes

Year	Authors (et al.)	Systems engineering approaches for RO membranes
1965	Lonsdale	Homogeneous diffusion model for cellulose acetate (CA) membrane [96]
1969	Hatfield	Nonlinear program for maximal flux and optimal arrangement of RO systems in brackish water [97]
1980	Tweddle	Prediction of performance of membrane modules with system analysis [98]
1982	Sirkar	Analytical design to estimate the averaged permeate solute concentration in spiral-wound RO module [99]
1984	van Dijk	Optimal design of total unit water cost using raw water TDS, pressure, and recovery [100]
1985	Evangellsia	Graphical-analytical method to design straight-through and tapered reverse osmosis plants [101]
1992	El-Halwagi	Optimal arrangement, types, and sizes of RO units for reverse osmosis networks (RON) [87]
1993	Sekino	Analytical model of friction–concentration–polarization (FCP) in hollow fiber RO module [102]
1996	Malek	Minimal cost analysis per unit membrane area applying large-sized permeates [41]
	Robertson	Dynamic matrix simulations for control of RO desalination pilot plant [103]
	Voros	Mathematical models for the performance of various SWRO process units [90]
	Sekino	Analytical model of FCP with Kimura–Sourirajan algorithm applied to hollow fiber RO module [104]

Desalination Technology for Sustainable Water Resource. Table 9 (Continued)

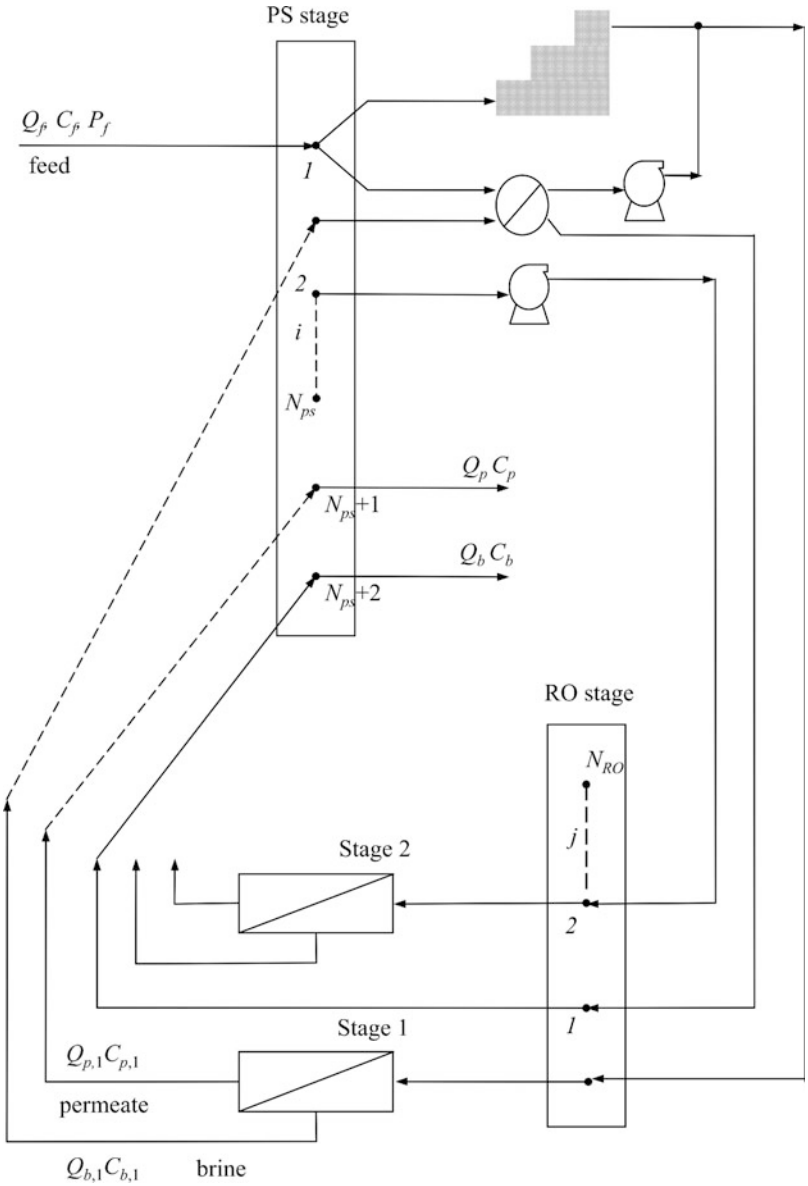
Year	Authors (et al.)	Systems engineering approaches for RO membranes
1997	Zhu	Optimal design of flexible RON with mixed-integer nonlinear programming (MINLP) [91]
	Voros	Optimal design to minimize total cost of RON plant [93]
1998	van der Meer	Hydraulic model of the rejection of mono- and bivalent ions in spiral-wound RO modules [105]
1999	Al-Bastaki	Mass transport model to predict the performance of hollow fiber RO membranes [106]
	See	RON desalination plant cost analysis with MINLP based on optimal cleaning schedule [92]
2000	Maskan	Optimization of RON operating conditions with a constrained multivariable nonlinear algorithm [88]
	Al-Bastaki	Mathematical analysis of concentration polarization and pressure drop based on flux integration [107]
2001	Wilf	Economic feasibility analysis of SWRO systems based on recovery rate and feedwater salinity [108]
2003	Al-Enezi	Design calculations based on feed salinity and temperature in RO desalination process [36]
	Villafafila	Optimization of operating and design parameters using successive quadratic programming (SQP) [109]
	Helal	Optimization of minimum water cost in hybrid RO/MSF system [110]
2004	Chatterjee	Numerical analysis of hollow fiber RO module using three-parameter Spiegler–Kedem (S–K) model [111]
2005	Marcovecchio	Minimization of total cost of hollow fiber RO seawater desalination using Kimura–Sourirajan model [112]

Desalination Technology for Sustainable Water Resource. Table 9 (Continued)

Year	Authors (et al.)	Systems engineering approaches for RO membranes
	Abbas	Feedforward neural network (NN) model to predict performance of RO experimental setup [113]
	Vitor Geraldes	Longitudinal variation model for mass/momentum transport in the spiral-wound SWRO modules [114]
2006	Senthilmurugan	Mathematical model for the separation of two solutes from aqueous solutions in hollow fiber module [115]
	Lu	Optimum design of SWRO system considering membrane cleaning and replacing based on MINLP [94]
2007	Lu	Optimum design of SWRO system under different feed concentration and product specification [95]

Cost Analysis Models

- *Total Cost (CAPEX/OPEX)*
  The total cost of a desalination plant can be divided into two terms: capital costs and operation and maintenance (O&M) costs [41, 58]. The capital cost can be considered as the total expenditures of the plant construction, which is the largest portion of the capital cost (50–85%), other installation costs (equipment, piping, service utilities, etc.), engineering efforts, and administrative/financing activities. O&M costs consist of plant operation costs (energy, chemicals, replacement of consumables, and labor) and maintenance costs for plant equipment, buildings, and utilities. The O&M costs are generally reported as all operational expenditures per year ($/year) or as operational costs for desalinated product water per volume ($/m^3).
- *Factors Influencing Total Costs*
  Figure 8 summarizes the factors affecting the total costs; these factors should be considered in order to obtain an accurate optimization.

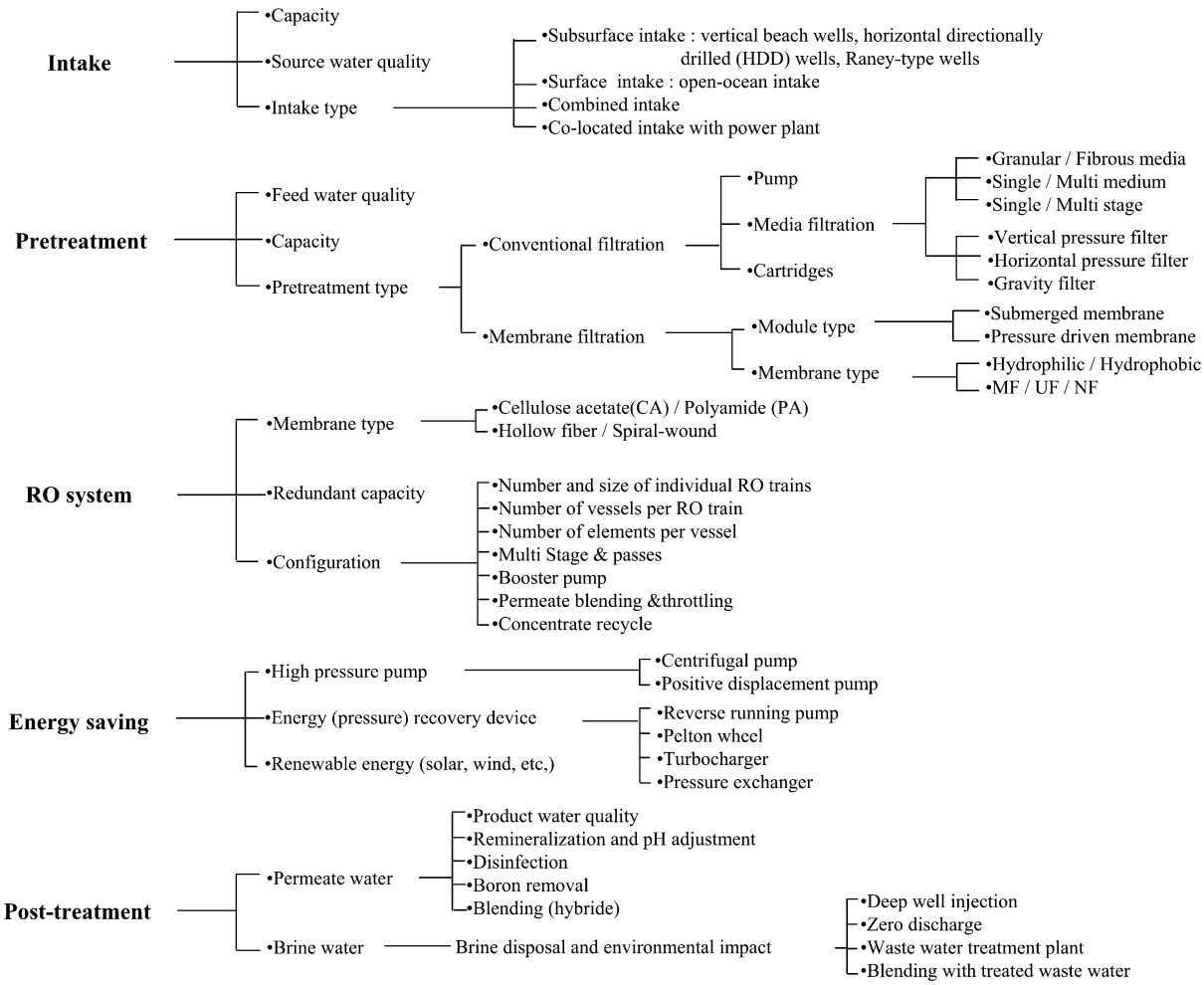


Desalination Technology for Sustainable Water Resource. Figure 7
Structural representation of an RO network (RON) [16]

- *Optimization of the Cost Model*
  There have been several models for estimating the membrane costs for seawater desalination plants released. For instance, the WTCost© Model was developed by the United States Bureau of Reclamation [116], and the Water Reuse Association in the US developed the WRA Model [117]. An example of the total cost optimization process that is commonly used in SWRO is given below [41, 91, 94, 95, 109, 112].

  A mixed-integer nonlinear programming (MINLP) model that minimizes the total cost can be solved by using an algebraic system. This model can be expressed as [94, 95]

$$J = \min_{\text{MINLP}} [C_T] \tag{13}$$



Desalination Technology for Sustainable Water Resource. Figure 8
System factors affecting the total cost in a large-scale SWRO plant [16]

where the objective function (J) is the total annualized cost (C_T).

$$C_T = (CC_{in} + CC_{hp} + CC_{px} + CC_{bp} + CC_m)1.411 \times 0.08 + OC_{in} + OC_{hp} + OC_{bp} + OC_m \quad (14)$$

In Eq. 14, CC and OC represent the annualized capital cost and annual operating cost, respectively; CC_{in}, CC_{hp}, CC_{px}, CC_{bp}, and CC_m represent the capital cost of the seawater intake pump, high-pressure pump, pressure exchanger, booster pump, and membrane purchase, and OC_{in}, OC_{hp}, and OC_{bp} indicate the energy costs for pump operations, and OC_m represents the maintenance cost for the membrane module, where 1.411 is the constant for the practice investment and 0.08 is the capital charge rate.

$$CC_{hp} = 52(\Delta P Q_{hp})^{0.96} \quad (15)$$

$$CC_{px} = 3134.7\ Q_{px}^{0.58} \quad (16)$$

$$Q_{ps,\,1} = Q_{hp} + Q_{px} \quad (17)$$

In Eqs. 15–17, P, Q_{hp}, and Q_{px} are the pressure and flow rates of the high-pressure pump and pressure exchanger, respectively; $Q_{ps,\,1}$ represents the flow rate of the first pressurization stage.

$$C_m = \sum_{j=1}^{N_{RO}} \sum_{k=1}^{8} Z_{j,k}\, C_k\, m_{j,k} n_j + \sum_{j=1}^{N_{RO}} C_{pv}\, n_j \quad (18)$$

In Eq. 18, C_m, C_k, and C_{pv} are the total cost of the membrane module, the price of the k_{th} membrane element, and the pressure vessel price, respectively. The indices j and k represent the j_{th} RO stage out of N_{RO} number of RO stages and the k_{th} membrane type out of a maximum of 8 elements. In addition, $Z_{j,\,k}$ is the binary variable, which is either 0 or 1; $m_{j,\,k}$ is the number of membrane elements in each pressure vessel; and n_j is the number of pressure vessels used in the j_{th} RO stage.

$$OC_{hp} = \frac{PQC_e f_c}{3.6\,\eta_{hp}\,\eta_{motor}} \tag{19}$$

$$\eta_{px} = \frac{\sum (PQ)_{out}}{\sum (PQ)_{in}} \times 100\% \tag{20}$$

In Eqs. 19 and 20, C_e is the electricity cost, and f_c is the load factor, where η_{hp}, η_{motor}, and η_{px} are the efficiencies of the high-pressure pumps, the electric motor, and PX, respectively.

Advantages and Limitations

RO membrane processes have a number of key advantages, such as a lower power requirement than other forms of desalination processes. They also require a smaller area for the RO membrane unit [50]; furthermore, RO processes can produce a high quality of water (i.e., TDS concentration between 100 and 500 ppm) [118], and can be operated automatically with minimal operator attention required [119].

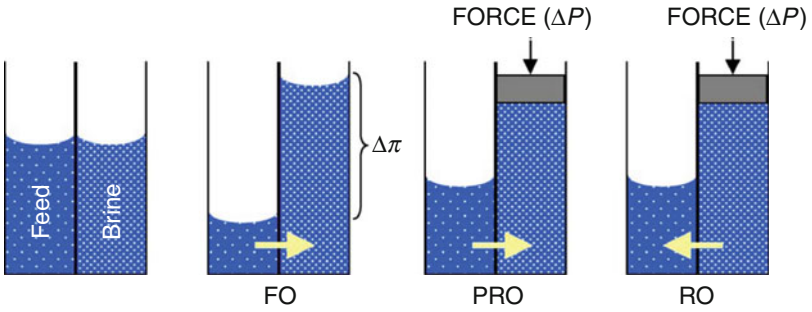
However, fouling and scaling are crucial limitations of RO membrane processes [118]. Fouling and scaling of the RO membrane reduce the quantity and quality of the water product. These phenomena can also lead to an increase in the water production costs due to their requirement for chemical cleanings. Good pretreatment practices are also required in order to maintain stable performance [118], and RO membranes are sensitive to oxidizing compounds (e.g., chlorine oxides); thus, feedwater needs to be treated in order to prolong the lifetime of the membrane [118].

Forward Osmosis (FO)

Principle

Forward osmosis (FO) , often described as "engineered osmosis" or an "osmotically driven process," has been spotlighted as a promising water treatment and desalination process. FO is a membrane-based separation process that utilizes natural osmosis phenomena. In this process, water can be transported across a semipermeable membrane by osmosis [120].

Water permeates through a semipermeable membrane from a low-concentration region to high-concentration region in the FO process (Fig. 9) [121], an exact opposite to the RO process. This water movement can be explained based on the chemical potential difference across the membrane; the difference in the chemical potential on the two sides of the membrane leads the solvent (pure water) to move from a higher potential environment (generally a lower solute concentration) to a lower potential environment (higher solute concentration) until they reach an equilibrium [122]. As a practical term, instead of "chemical

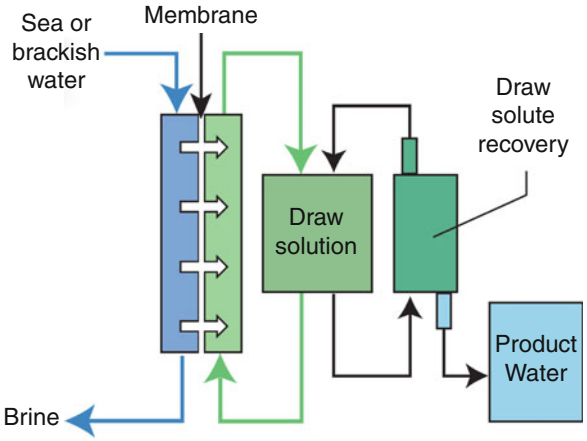


Desalination Technology for Sustainable Water Resource. Figure 9
Classification of membrane processes [121]. FO is naturally driven by osmotic pressure. If the applied pressure exceeds the osmotic pressure of the draw solution, the process is called RO. Pressure-retarded osmosis (PRO) refers to the case when the applied pressure is lower than the osmotic pressure

potential," osmotic pressure is used here to refer to the pressure that can induce the spontaneous net flow of water across a semipermeable membrane. Therefore, the driving force of this process is created by the transmembrane osmotic pressure difference. The higher concentration solution, exerting a higher osmotic pressure, is referred to as the draw solution.

System Investigation

Figure 10 presents the conceptual diagram of the FO process, which mainly consists of two major stages: (1) a membrane process and (2) the separation and recovery of the draw solution. During the membrane process, pure water spontaneously flows into the draw solution side due to the osmotic pressure difference. While the feedwater is concentrated and finally discarded as brine, the draw solution is simultaneously diluted with the permeating water. The diluted draw solution is then sent to the separation and recovery unit for the next step of the FO process. Next, the product water is separated from the diluted draw solution to yield potable water for several consumption purposes. Simultaneously, the draw solute is recovered and recycled back to the membrane process [124]. The currently suggested separation and recovery processes include evaporation/condensation, RO, distillation, membrane distillation (MD), and magnetic separation [125]. Note that the product water quality and draw solute recovery rate are the most important parameters for FO process as they reveal whether the process is feasible or not. During this closed-loop continuous process, loss of the draw solute should be minimized in order to lower its replenishment cost [126]. The draw solute can possibly be lost due to reverse diffusion during the membrane process and insufficient separation of the draw solute within the recovery unit [127].



Desalination Technology for Sustainable Water Resource. Figure 10
Forward osmosis desalination process schematic adopted from McCutcheon et al. [123]

In addition to its water treatment/production applications, FO has been increasingly addressed for energy generation. FO can potentially be applied to generate energy by utilizing the concentration gradient between seawater and freshwater (Fig. 11) [126]. This process is referred to as pressure-retarded osmosis (PRO). In PRO, transmembrane water permeation pressurizes the draw solution side and power is harvested via depressurization through a hydro turbine. In 2009, Statkraft, located in Norway, opened the first prototype PRO installation; the plant configuration followed the original schematic proposed by Sidney Loeb.

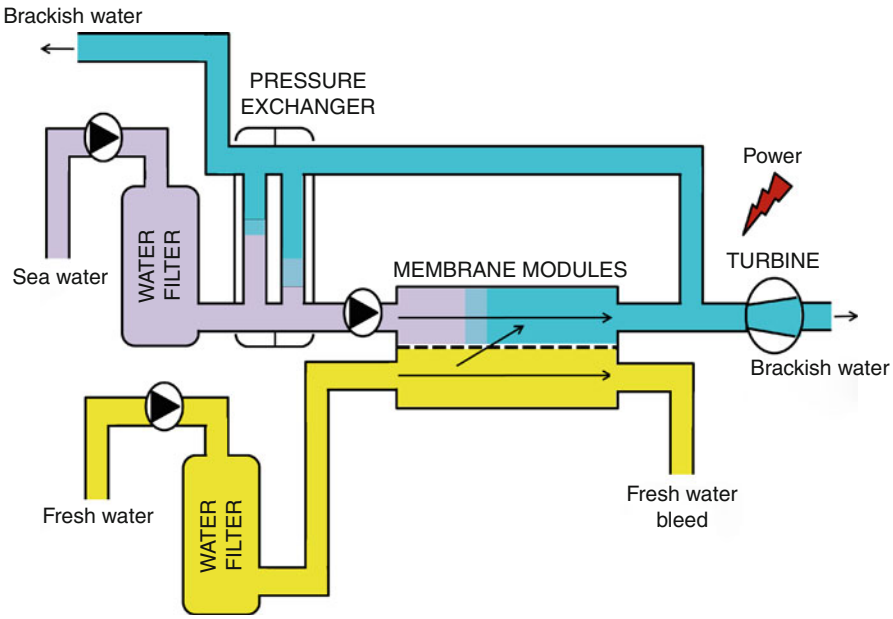
System Model

In order to predict the flux performance in FO, a standard flux equation can be applied [123]:

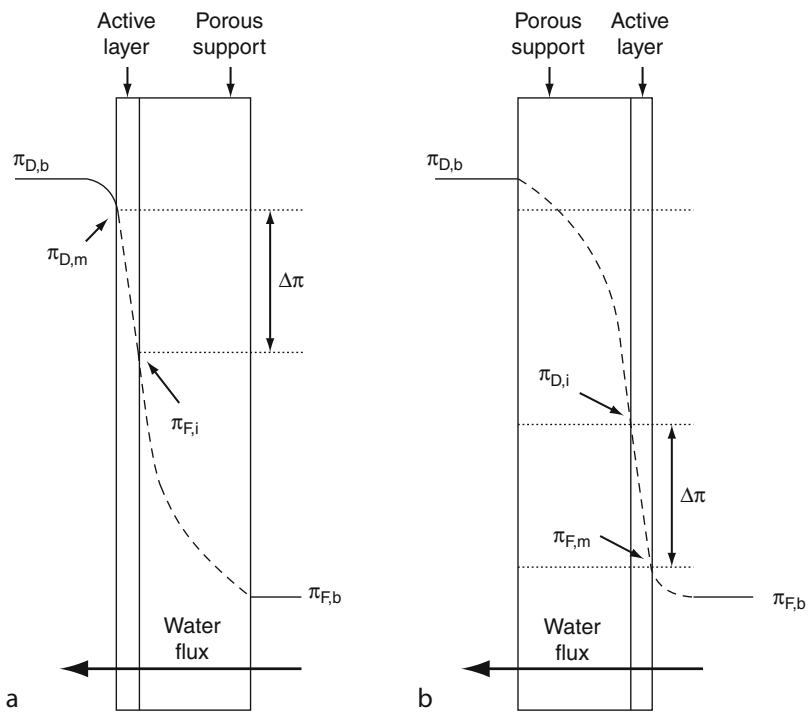
$$J_w = A(\pi_{D,b} - \pi_{F,b}) \tag{21}$$

where A is the water permeability, and $\pi_{D,b}$ and $\pi_{F,b}$ are the bulk osmotic pressures at the feed and draw sides, respectively. This equation is representation of an FO process that is driven by the osmotic pressure difference. However, it has been found that actual membrane experiments have resulted in much less flux performance than was calculated from the bulk osmotic pressure difference [123]. These unexpected results were caused by a phenomenon referred to as the concentration polarization (CP). CP can be basically explained as the increased and/or decreased concentration on the membrane surface that thereby lowers the net osmotic pressure difference. Therefore, the osmotic pressure terms used in the standard flux equation should be modified by taking CP into account. Membrane orientation is an important factor for CP calculations; since the membranes normally adopted for FO are asymmetric, the position of the active layer also plays an important role.

The concentrations on both sides of the membrane affect the flux performance in FO, whereas RO only takes



Desalination Technology for Sustainable Water Resource. Figure 11
Simplified process layout for a PRO power plant [121]



Desalination Technology for Sustainable Water Resource. Figure 12
Illustration of osmotic pressure profile in an asymmetric membrane [123]

the feed side concentration into account. As described in Fig. 12 [123], the actual driving force ($\Delta\pi$) is dramatically reduced by four types of CP. In PRO mode, dilutive external concentration polarization (ECP) at the left side of the membrane and concentrative internal concentration polarization (ICP) within the porous support layer can be observed. In FO mode, dilutive ICP inside the porous support layer and concentrative ECP on the active dense layer surface can be observed.

Based on the following equation, the ECP can be analyzed in a similar way as in RO [72]:

$$\frac{c_w - c_p}{c_0 - c_p} = \exp\left(\frac{\bar{v}}{k}\right) \tag{22}$$

where c_w, c_p, and c_0 are the solute concentrations on the membrane surface, in the permeate, and bulk solutions, respectively, $\bar{v}$ is the channel averaged permeate velocity, and k is the mass transfer coefficient. In order to model the ECP, the mass transfer coefficient (defined as $k = D/\delta$, where D is the diffusion coefficient of the solute and δ is the concentration boundary layer thickness) can be acquired based on the Sherwood relationship [72]:

$$\mathrm{Sh} = a\mathrm{Re}^b\mathrm{Sc}^c \tag{23}$$

If it is assumed that the membrane completely rejects the solutes and that the osmotic pressure and concentration are linearly proportional, the ECP modulus for both PRO and FO modes can then be given by the modification of Eq. 22 [123], such that

$$\frac{\pi_{D,m}}{\pi_{D,b}} = \exp\left(-\frac{J_w}{k}\right) \tag{24}$$

$$\frac{\pi_{F,m}}{\pi_{F,b}} = \exp\left(\frac{J_w}{k}\right) \tag{25}$$

where $\pi_{D,m}$ and $\pi_{D,b}$ refer to the osmotic pressures of the draw solution on the membrane surface and in the bulk solution, respectively, and $\pi_{F,m}$ and $\pi_{F,b}$ are the osmotic pressures of the feed solution on the membrane surface and in the bulk solution. Equation 24 defines the dilutive ECP and Eq. 25 is for the concentrative ECP.

The ICP phenomenon, particularly observed during mass transport, is the major factor that lowers the water flux in an FO process. The CP, which occurs inside a porous support layer, cannot be removed by physical methods such as increasing the cross flow velocity. Indeed, ICP originates from the characteristics of a symmetric membrane, which has a porous support layer that acts like an undisturbed concentration boundary layer.

Lee et al. [128] proposed a flux model for PRO based on a consideration of the ICP phenomenon. This model can be derived by applying the solute mass balance within the porous support layer to the boundary conditions:

$$B(C_{D,m} - C_{F,i}) = D\varepsilon\frac{dC(x)}{ds} - J_w C(x) \tag{26}$$

$$\begin{aligned} C(x) &= C_{F,i} \quad at\ x = 0 \\ C(x) &= C_{F,b} \quad at\ x = t\tau \end{aligned} \tag{27}$$

where B is solute permeability, ε is the porosity of the porous support layer, and t and τ are the thickness and tortuosity of the porous support layer, respectively ($t\tau$ corresponds to the distance of the effective boundary layer porous layer).

The following equation can then be obtained as a result of the integration of Eq. 26:

$$\frac{C_{F,i}}{C_{D,m}} = \frac{B[\exp(J_W K - 1) + J_W \frac{C_{F,b}}{C_{D,m}}\exp(J_w K)}{B[\exp(J_W K) - 1] + J_W} \tag{28}$$

where K is the solute resistivity (defined as $K = t\tau/D\varepsilon$), which describes how resistant the solute is to the porous layer. In order to simplify the equation, a linear relationship between the concentration and osmotic pressure is assumed. Therefore, the flux equation for PRO can be rewritten as

$$J_w = A\frac{1 - \frac{C_{F,b}}{C_{D,m}}\exp(J_w K)}{1 + \frac{B}{J_w}[\exp(J_W K) - 1]} \tag{29}$$

Loeb [128] further simplified Eq. 29 by applying the linear proportionality between the concentration and the osmotic pressure in order to obtain the solute resistivity:

$$K = \frac{1}{J_w}\left(\ln\frac{B + A\pi_{D,m} - J_w}{B + A\pi_{F,b}}\right) \text{ PRO mode} \tag{30}$$

$$K = \frac{1}{J_w}\left(\ln\frac{B + A\pi_{D,b}}{B + J_w + A\pi_{F,m}}\right) \text{ FO mode} \tag{31}$$

The ICP modulus can then be obtained from the above equations when perfect solute rejection ($B = 0$) is assumed:

$$\frac{\pi_{F,i}}{\pi_{F,b}} = \exp(J_w K) \tag{32}$$

$$\frac{\pi_{D,i}}{\pi_{D,b}} = \exp(-J_w K) \tag{33}$$

Equation 32 can define the concentrative ICP, whereas Eq. 33 is for the dilutive ICP. The modification of the osmotic pressure terms in Eq. 21 with ECP and ICP modulus gives the following final flux equations for the PRO and FO modes [123]:

$$J_w = A[\pi_{D,b}\exp(-\frac{J_w}{k}) - \pi_{F,b}\exp(-J_w K)]\ \text{PRO mode} \tag{34}$$

$$J_w = A[\pi_{D,b}\exp(-J_w K) - \pi_{F,b}\exp(\frac{J_w}{k})]\ \text{FO mode} \tag{35}$$

In another study, the effects of system parameters on FO membranes were simulated using a mathematical process model [129]. This model used CP for the plate and frame (PNF) module. Membrane orientation, flow direction, flow rate, and solute resistivity were selected as system parameters. It was found that the flow directions of the feed and draw solutions have no significant influence on the FO performance. In the case of membrane orientation, the all-inside case, in which the draw solution faces the active layer, displays a relatively higher performance than all-outside and all-up cases. Notably, the membrane performances were found to be affected by K, indicating the extent of the internal CP.

In order to investigate the importance of the membrane structure in FO process, the membrane structural parameter (S) was defined [130]; S was defined as being independent of the draw solution properties, assuming that osmotically active solutes do not physically affect the membrane layers. As such, S can be used to determine the degree of ICP in the porous support structure of an FO membrane [131].

$$S = KD = \frac{t_S \tau}{\varepsilon} \tag{36}$$

where K is the membrane support layer resistance to solute diffusion, D is the diffusion coefficient of the draw solute, t_s is the support layer thickness, τ the tortuosity, and ε the porosity. The intrinsic water permeability (A) and NaCl permeability coefficient (B), which are used to calculate the K value, were found to have an intricate interrelationship with S. Therefore A, B, and S were suggested to be tailored in order to improve the membrane performance.

The most influential factor causing inconsistent S values was found to be the assumption of identical concentration and osmotic pressure ratios, which are frequently used in FO modeling. Therefore, an FEM-based model was developed with the aim of obtaining a constant S parameter [131]. This developed model could successfully simulate the FO performance with a constant S value under different feed and draw concentrations.

Advantages and Limitations

Theoretically, FO processes have a low energy consumption rate because they do not require hydraulic pressure as the driving force. RO, on the other hand, requires a hydraulic pressure of at least more than twice the osmotic pressure of seawater (~25 bar) [132]. In addition, without hydraulic pressure, the fouling behavior of FO processes is significantly different than in RO. Fouling is relatively reversible in FO processes by physical cleaning because the fouling layer is less compact and sparser than that observed in RO [133]. The recovery rate in FO may also be higher than in other membrane processes and thus can result in less brine discharge to the environment [134]. Therefore, FO processes have significant potential advantages over pressure-driven membrane processes such as RO, NF, etc.

However, even though FO has a number of advantages over pressure-driven processes, the practical application of FO is still under investigation. Two major problems of FO processes that need to be resolved are the lack of suitable FO membranes and the development of separation and recovery technologies.

In order to enhance the flux performance in FO processes, appropriate membranes that can minimize ICP phenomenon should be developed since ICP is the major factor for flux decline. A suitable semipermeable membrane must have high water permeability

(a porous layer with less thickness and higher porosity) and a high solute rejection rate [135].

Challenges pertaining to the selection of a suitable draw solution also remain as priority; complete separation of the draw solute should be guaranteed for drinking water production [136]. In addition, the draw solute should be easily recovered by reconcentration or reconstitution in order for it to be recirculated in a closed-loop system without loss. Establishment of a stable technology for this separation and recovery process can enable continuous water production. For this task, draw solutions with high osmotic pressure, a relatively big molecular size, suitable separation properties (boiling point, vapor pressure, etc.), and high water solubility are recommended.

Membrane Distillation (MD)

Membrane distillation (MD) has been studied worldwide as an attractive membrane separation process. Since it was first introduced in 1960s, there have been key improvements and developments in the system, such as for membrane engineering [137–139]. Figure 13 presents the volume of related studies published in this field. The current level of interest in MD technology started from the late 1990s; however, further studies and developments are required for prior to actual industrial applications of the technology [140–142].

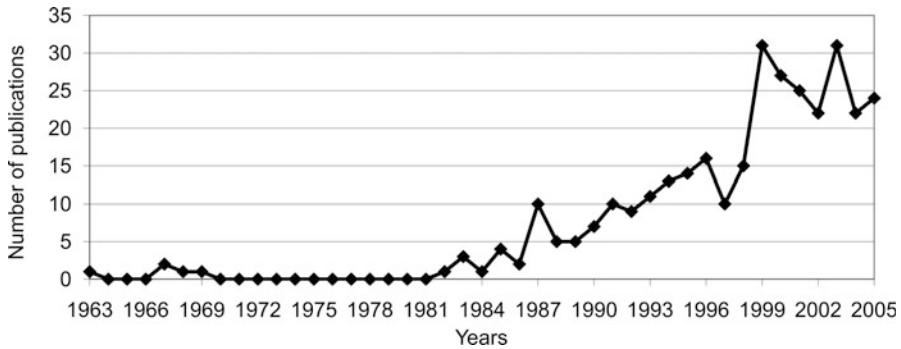
Principle

MD is a promising alternative technology to conventional separation processes such as RO and thermal distillation, with lower cost and energy requirements. MD can be defined as a thermal-driven membrane process, in which only molecular water in the form of steam can transport through microporous hydrophobic membranes [143]. In this process, the aqueous feed solution, which is in direct contact with the feed side of the membrane, cannot flow through the membrane pores because of its hydrophobic nature. Consequently, interfaces between the liquid and vapor phases are formed at the membrane pore entrance, such that the driving force of MD is the vapor pressure difference across the membrane.

System Investigation

In order to create and maintain the driving force through the membrane, there are four types of configurations for MD technology:

1. Direct contact membrane distillation (DCMD): the permeate side of the membrane is in direct contact with the liquid phase, which is colder than the feed solution.
2. Air gap membrane distillation (AGMD): an air gap is interposed between the membrane and the condensing surface.
3. Sweep gas membrane distillation (SGMD): cold inert gas is used for sweeping the permeate side of the membrane to carry the water vapor molecules.
4. Vacuum membrane distillation (VMD): a vacuum is used on the permeate side of the membrane in order to increase the driving force.



Desalination Technology for Sustainable Water Resource. Figure 13
Level of interest in MD processes [141]

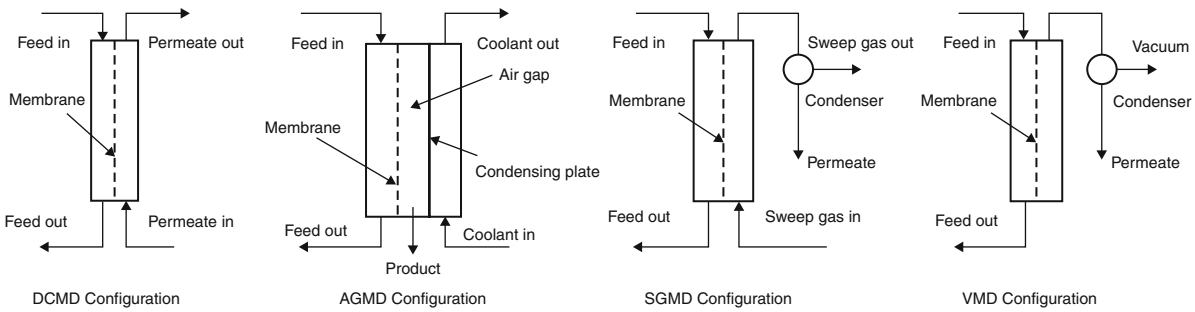
Among these configurations, DCMD and AGMD are considered the most suitable for desalination technology [144–150].

System Model

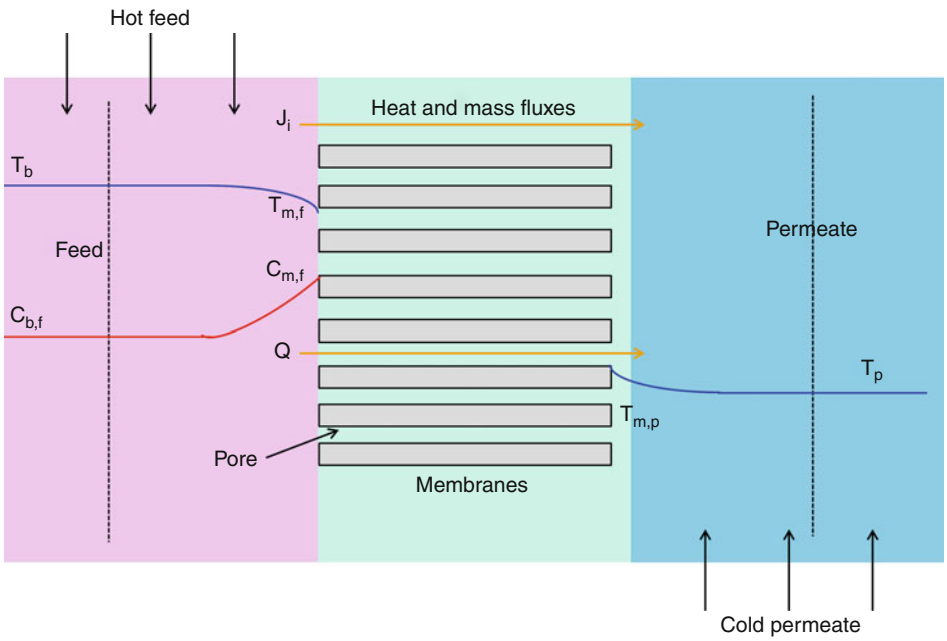
Two of the main functions of an MD model are to predict the permeate flux and to estimate the temperature and concentration polarization coefficients, which are dependent on the membrane module design, operating variables, and membrane parameters. Furthermore, membrane selectivity can also be predicted for VMD configurations (Fig. 14).

As such, the effects of both temperature and concentration polarizations should be considered since heat and mass transfers occur simultaneously in MD processes (Figs. 15 and 16).

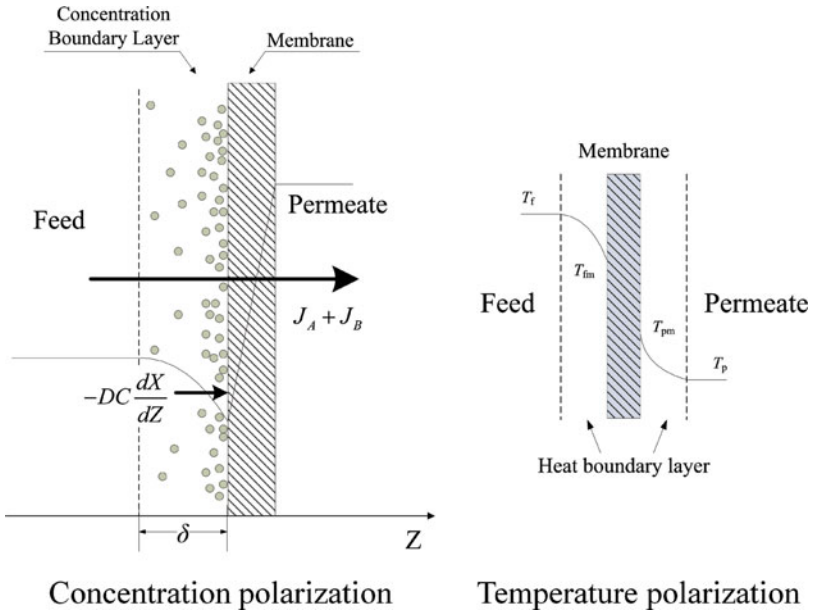
These phenomena are typically analyzed based on the assumption that vapor permeates through a microporous membrane. This flow of vapor consists of three contributions: the Knudsen flow, Poiseuille



Desalination Technology for Sustainable Water Resource. Figure 14
MD configurations [141]



Desalination Technology for Sustainable Water Resource. Figure 15
Heat and mass transfer through the membrane in a DCMD configuration [151]



Desalination Technology for Sustainable Water Resource. Figure 16
Phenomena of temperature and concentration polarization

flow, molecular diffusion flow, and transitions between them. The dusty gas model is usually used to calculate the MD fluxes, and the molecular diffusion flow is not applied in VMD systems. Furthermore, numerical analyses have been developed by using momentum, energy, and diffusion equations.

Heat Transfer Heat transfer consists of four steps:

1. Heat transfer from the feed to the membrane surface across the thermal boundary layer on the feed side of the membrane, which can be associated with the temperature polarization effect
2. Heat transport by conduction, which is generated across the membrane matrix and heat loss occurs in the vapor filled pores
3. Mass transfer through the membrane pores as a result of the latent heat of vaporization
4. Heat transfer from the membrane surface to the permeate solution across the thermal boundary layer, which can be related to the temperature polarization effect on the feed side

In order to calculate the temperature polarization coefficient (TPC), which is defined as the fraction of difference between the transmembrane temperature and bulk temperature, the following equation is used:

$$TPC = \frac{T_{fm} - T_{pm}}{T_{fb} - T_{pb}} \tag{36}$$

where T_{fm}, T_{pm}, T_{fb}, T_{pb} are temperatures of the membrane surface and bulk on the feed and permeate sides. The latent heat for vaporization should be continuously transferred from the feed to the membrane surface because MD depends on the phase change. Thus, the equation of heat flux can be expressed as

$$\mathrm{Q_f} = \mathrm{h_f}(\mathrm{T_{fb}} - \mathrm{T_{fm}}) \tag{37}$$

where $\mathrm{Q_f}$ is the heat flux, which relies on the film heat transfer coefficient, and $\mathrm{h_f}$ is the temperature difference between the feed and membrane surface.

However, it should be noted that even though a higher feed temperature brings an exponential increase in the flux across the membrane, the effect of temperature polarization also increases with the feed temperature [148, 150].

Mass Transfer The phenomena of mass transport of volatile molecules can be explained as follows:

1. It takes place from the bulk feed to the membrane surface.
2. The vapor molecules are transported through the membrane pores.
3. It also takes place from the membrane surface to the bulk permeate phase on the permeate side.

The molar flux related to mass transport through the membrane pores can then be defined as

$$N = C\left[P_{\mathrm{fm}}(T_{\mathrm{fm}}) - P_{\mathrm{pm}}(T_{\mathrm{pm}})\right] \tag{38}$$

where C is the membrane distillation coefficient and $[P_{\mathrm{fm}}(T_{\mathrm{fm}}) - P_{\mathrm{pm}}(T_{\mathrm{pm}})]$ is the vapor pressure difference across the membrane [141].

Advantages and Limitations

Membrane distillation requires a low external energy for operation and has a relatively low capital cost for plant installation. In contrast to conventional distillation processes that require a large vapor space and high vapor velocities to provide vapor/liquid contact, MD uses a hydrophobic microporous membrane to create a vapor/liquid interface. Consequently, the MD module is much smaller than the conventional processes, which means it can minimize the capital cost and operate at much lower temperatures. In addition, the lower temperature with reduced equipment surface area can significantly decrease the heat loss [152, 153]. Therefore, alternative energy sources such as geothermal and solar energies can be applied in MD systems. Key advantages of the MD process are summarized below [143]:

- Hundred percent (theoretical) rejection of ions, macromolecules, colloids, and other nonvolatiles
- Reduced vapor spaces and lower operating temperatures compared to traditional distillation processes
- Lower operating pressures compared to conventional pressure-driven membrane separation processes
- Less chemical interaction between membrane and process solutions
- Reduction in required membrane mechanical properties

Although MD has notable advantages over other processes, a number of limitations for commercial applications in water desalination remain; these include [141]:

- A relatively low permeate flux in comparison with other separation process, such as RO
- Permeate flux decline because of concentration and temperature polarization effects, membrane fouling, and pore wetting
- A lack of membrane and module designs for MD
- High thermal energy consumption

Future Directions

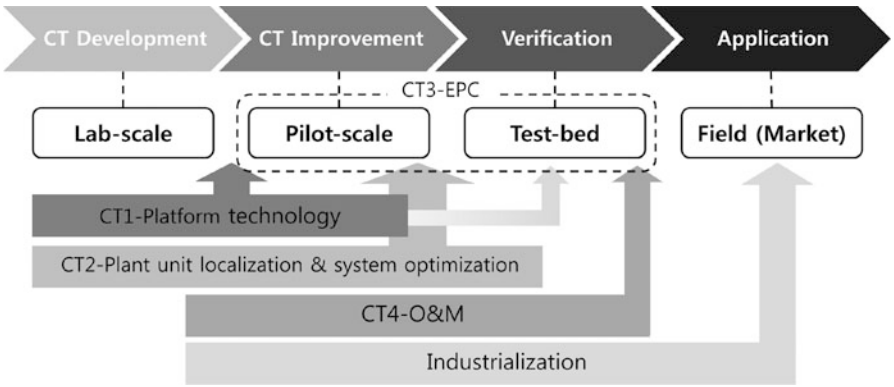
Thermal Process Thermal desalination technologies, including MSF and MED, require an energy-intensive process; therefore, most research has focused on performance improvements and design simplifications.

- Development of alternative energy sources
- Mitigation and control of scaling and fouling
- Prevention and control of scaling and corrosion
- Alternative construction materials
- Optimization of process design
- Improvements in component design
- Control systems for optimizing consumables depletion

RO Process It is important to continue further development and investment efforts in SWRO desalination programs to resolve water scarcity problems in regions around the world, with the ultimate goal of reducing the cost of final water production. Subsequently, several researches to make the cost of product water more economical and enhance the process performance are suggested. These include:

- Optimization of pretreatment process design
- Development of membrane performance (durability, capacity, efficiency)
- Application of renewable energy resources
- Investigation of brine disposal (aimed for zero discharge)

There have been a number of new SWRO centers established, most by governments and big corporations, in order to promote and improve SWRO desalination. For example, the Seawater Engineering and Architecture of High Efficiency Reverse Osmosis (SeaHERO) center is one of the most important and



Desalination Technology for Sustainable Water Resource. Figure 17
SeaHERO's R&D strategy based on 4 CT projects and test-bed [154]

impressive centers, with a US $138.7 million budget, and involving 753 researchers from 50 different organizations. The R&D strategy of SeaHERO is based on test-bed-related core technology (CT) developments (Fig. 17).

In SeaHERO, there are four CT projects, including development of platform technologies for SWRO plant construction (CT 1: Platform technology), development of SWRO membranes and high-pressure pump component manufacturing and system optimization technologies (CT 2: Plant unit localization and system optimization), development of large-scale SWRO plant design and construction technology (CT 3: Engineering–procurement–construction (EPC)), and development of innovative operation and management (O&M) technology for large-scale SWRO plants (CT 4: O&M) [154, 155].

FO Process The ongoing studies related to FO are still at the lab-scale stage and have mainly focused on the membrane process. Therefore, FO is not currently available for water treatment applications as a standalone process, and will not be until these limitations can be resolved. This means that even though FO processes have been presumed to have great advantages, their practical application in real treatment plants is still far from reality, unless a reliable draw solution can be developed. Indeed, the forecast of low energy consumption, one of the biggest merits of FO, is theoretically based solely on the membrane process without considering the effective separation and recovery step. In order to be accepted as an economically feasible process, the development of membrane exclusively for FO should be progressed carefully, connected to advances in separation and recovery technologies. For the meantime, FO is expected to be hybridized with other processes, such as RO and MBR, since they have more feasible application possibilities for wastewater treatment or water reuse than for desalination.

MD Process There are several barriers that need to be overcome prior to the industrial implementation of MD processes, including [140, 141]:

- A lack of understanding pertaining to temperature and concentration polarizations, and surface membrane fouling.
- A lack of understanding of heat transfer in the MD process.

Therefore, these problems have to be investigated with respect to cost analysis, long-term operation, membrane modules, membrane fouling, and development of transfer models.

As a possible solution, in order to optimize energy requirements, hybrid systems such as RO/MD and NF/RO/MD can be offered [156]. However, although the energy consumption associated with coupling systems are higher than those of RO alone, the overall performance can be more efficient [157]. Thus, these integrated MD systems may yet be competitive alternatives to RO if thermal energy becomes readily available.

Bibliography

1. UNDP (2006) Human Development Report 2006
2. WHO (2005) The World Health Report 2005
3. Global Water Intelligence (2010) Global water market 2011, vol 1, pp 101–121, Media Analytics Ltd.
4. Khawaji A, Kutubkhanah I, Wie J (2008) Advances in seawater desalination technologies. Desalination 221:47–69
5. Global Water Intelligence (2011) DesalData.com
6. Abdul-Wahab SA, Reddy KV, Al-Weshahi MA, Al-Hatmi S, Tajeldin YM (2011) Development of a steady-state mathematical model for multistage flash (MSF) desalination plant. Int J Energy Res (doi:10.1002/er.1826)
7. Hussain AA, Nataraj SK, Abashar MEE, Al-Mutaz IS, Aminabhavi TM (2008) Prediction of physical properties of nanofiltration membranes using experiment and theoretical models. J Membr Sci 310:321–336
8. Zhao D, Xue J, Li S, Sun H, Zhang Q-D (2011) Theoretical analyses of thermal and economical aspects of multi-effect distillation desalination dealing with high-salinity wastewater. Desalination 273:292–298
9. Engelien HK, Larsson T, Skogestad S (2001) Simulation and optimization of heat integrated distillation columns. In: 42nd SIMS conference, Porsgrunn, Norway, pp 367–376
10. USEPA (2005) Membrane filtration guidance manual United States Environmental Protection Agency, EPA, 815-R-06-009
11. Van der Meer WGJ (2003) Mathematical modeling of NF and RO membrane filtration plants and modules. Ph.D. Thesis, Delft University of Technology, Netherlands
12. DOW filmtec (2005) Technical manual for FFILMTEC reverse osmosis membranes: liquid separation. Dow filmtec, Beulah, Michigan
13. Baker RW (2004) Membrane technology and applications, 2nd edn. Wiley, West Sussex
14. Wilf M (2007) The guidebook to membrane desalination technology: reverse osmosis, nanofiltration and hybrid systems, process, design, applications and economics. Balaban Desalination Publication, L'Aquila, Italy
15. Bhattacharyya D, Williams M (1992) Theory – reverse osmosis. Van Nostrand Reinhold, New York
16. Kim YM, Kim SJ, Kim YS, Lee S, Kim IS, Kim JH (2009) Overview of systems engineering approaches for a large-scale seawater desalination plant with a reverse osmosis network. Desalination 238:312–332
17. Eriksson P (1988) Nanofiltration extends the range of membrane filtration. Environ Prog 7:58–62
18. Van der Bruggen B, Vandecasteele C (2003) Removal of pollutants from surface water and groundwater by nanofiltration: overview of possible applications in the drinking water industry. Environ Pollut 122:435–445
19. Yu S, Gao C, Su H, Liu M (2001) Nanofiltration used for desalination and concentration in dye production. Desalination 140:97–100
20. Amy G (2007) Membrane-based water desalination – state of the art and future prospects. In: International seminar by center for seawater desalination plant, Seoul, Korea
21. El-Manharawy S, Hafez A (2001) Water type and guidelines for RO system design. Desalination 139:97–113
22. Al-Wazzan Y, Safar M, Ebrahim S, Burney N, Mesri A (2002) Desalting of subsurface water using spiral-wound reverse osmosis (RO) system: technical and economic assessment. Desalination 143:21–28
23. Ebrahim S, Al-Wazzan Y, Safar M, Burney N, Al-Mesri A (2000) Pilot study on renovation of subsurface water using a reverse osmosis desalting system. Desalination 131:315–324
24. El-Azizi IM, Omran AAM (2003) Design criteria of 10,000 m^3/d SWRO desalination plant of Tajura, Libya. Desalination 153:273–279
25. Elguera AM, Pérez Báez SO (2005) Development of the most adequate pre-treatment for high capacity seawater desalination plants with open intake. Desalination 184: 173–183
26. Khawaji AD, Kutubkhanah IK, Wie J-M (2007) A 13.3 MGD seawater RO desalination plant for Yanbu industrial city. Desalination 203:176–188
27. Leparc J, Rapenne S, Courties C, Lebaron P, Croué JP, Jacquemet V, Turner G (2007) Water quality and performance evaluation at seawater reverse osmosis plants through the use of advanced analytical tools. Desalination 203:243–255
28. Muirhead A, Beardsley S, Aboudiwan J (1982) Performance of the 12,000 m^3/d seawater reverse osmosis desalination plant at Jeddah, Saudi Arabia January 1979 through January 1981. Desalination 42:115–128
29. Sadhwani JJ, Veza JM, Santana C (2005) Case studies on environmental impact of seawater desalination. Desalination 185:1–8
30. Teuler A, Glucina K, Laîné JM (1999) Assessment of UF pretreatment prior RO membranes for seawater desalination. Desalination 125:89–96
31. Zidouri H (2000) Desalination in Morocco and presentation of design and operation of the Laayoune seawater reverse osmosis plant. Desalination 131:137–145
32. Kim SL, Paul CJ, Ting YP (2002) Study on feed pretreatment for membrane filtration of secondary effluent. Sep Purif Technol 29:171–179
33. Sheikholeslami R, Al-Mutaz IS, Koo T, Young A (2001) Pretreatment and the effect of cations and anions on prevention of silica fouling. Desalination 139:83–95
34. Pearce GK (2007) The case for UF/MF pretreatment to RO in seawater applications. Desalination 203:286–295
35. Wolf PH, Siverns S, Monti S (2005) UF membranes for RO desalination pretreatment. Desalination 182:293–300
36. Al-Enezi G, Fawzi N (2003) Design consideration of RO units: case studies. Desalination 153:281–286
37. Cho J, Amy G, Pellegrino J (2000) Membrane filtration of natural organic matter: factors and mechanisms affecting rejection and flux decline with charged ultrafiltration (UF) membrane. J Membr Sci 164:89–110
38. Kim S, Hoek EMV (2007) Interactions controlling biopolymer fouling of reverse osmosis membranes. Desalination 202: 333–342

39. Luo M, Wang Z (2001) Complex fouling and cleaning-in-place of a reverse osmosis desalination system. Desalination 141:15–22
40. Madaeni SS, Mohamamdi T, Moghadam MK (2001) Chemical cleaning of reverse osmosis membranes. Desalination 134:77–82
41. Malek A, Hawlader MNA, Ho JC (1996) Design and economics of RO seawater desalination. Desalination 105:245–261
42. Matsuura T (2001) Progress in membrane science and technology for seawater desalination - a review. Desalination 134:47–54
43. Pohland HW (1981) Seawater desalination and reverse osmosis plant design. In: Coating conference, proceedings of the technical association of the pulp and paper industry, Atlanta, pp 157–167
44. Sablani SS, Goosen MFA, Al-Belushi R, Wilf M (2001) Concentration polarization in ultrafiltration and reverse osmosis: a critical review. Desalination 141:269–289
45. Sheikholeslami R, Tan S (1999) Effects of water quality on silica fouling of desalination plants. Desalination 126:267–280
46. Tay KG, Song L (2005) A more effective method for fouling characterization in a full-scale reverse osmosis process. Desalination 177:95–107
47. Yiantsios SG, Sioutopoulos D, Karabelas AJ (2005) Colloidal fouling of RO membranes: an overview of key issues and efforts to develop improved prediction techniques. Desalination 183:257–272
48. Zhou W, Song L, Guan TK (2006) A numerical study on concentration polarization and system performance of spiral wound RO membrane modules. J Membr Sci 271:38–46
49. Barger M, Carnahan RP (1991) Fouling prediction in reverse osmosis processes. Desalination 83:3–33
50. Fritzmann C, Lowenberg J, Wintgens T, Melin T (2007) State-of-the-art of reverse osmosis desalination. Desalination 216:1–76
51. Hoek EMV, Allred J, Knoell T, Jeong B (2008) Modeling the effects of fouling on full-scale reverse osmosis process. J Membr Sci 314:33–49
52. Beier SP (2006) Pressure driven membrane processes, Ventus Publishing, Copenhagen, Denmark
53. Hoek EMV, Kim AS, Elimelech M (2002) Influence of crossflow membrane filter geometry and shear rate on colloidal fouling in reverse osmosis and nanofiltration separations. Environ Eng Sci 19:357–372
54. Kimura S, Nakao S (1975) Fouling of cellulose acetate tubular reverse osmosis modules treating the industrial water in Tokyo. Desalination 17:267–288
55. Ng HY, Elimelech M (2004) Influence of colloidal fouling on rejection of trace organic contaminants by reverse osmosis. J Membr Sci 244:215–226
56. Shaalan HF (2002) Development of fouling control strategies pertinent to nanofiltration membranes. Desalination 153: 125–131
57. Kurihara M, Yamamura H, Nakanishi T, Jinno S (2001) Operation and reliability of very high-recovery seawater desalination technologies by brine conversion two-stage RO desalination system. Desalination 138:191–199
58. Avlonitis SA (2002) Operational water cost and productivity improvements for small-size RO desalination plants. Desalination 142:295–304
59. Stover RL (2004) Development of a fourth generation energy recovery device. A 'CTO's notebook'. Desalination 165:313–321
60. Wilf M, Schierach MK (2001) Improved performance and cost reduction of RO seawater systems using UF pretreatment. Desalination 135:61–68
61. Wilf M, Bartels C (2005) Optimization of seawater RO systems design. Desalination 173:1–12
62. Bick A, Oron G (2005) Post-treatment design of seawater reverse osmosis plants: boron removal technology selection for potable water production and environmental control. Desalination 178:233–246
63. Reynolds T, Debroux J (2007) Seawater desalination pilot program: First and second pass RO permeate and finished water quality, Marin Municipal Water District, Corte Madera, CA
64. Jacob C (2007) Seawater desalination: boron removal by ion exchange technology. Desalination 205:47–52
65. Nadav N (1999) Boron removal from seawater reverse osmosis permeate utilizing selective ion exchange resin. Desalination 124:131–135
66. Taniguchi M, Kurihara M, Kimura S (2001) Boron reduction performance of reverse osmosis seawater desalination process. J Membr Sci 183:259–267
67. Ahmed M, Shayya WH, Hoey D, Al-Handaly J (2001) Brine disposal from reverse osmosis desalination plants in Oman and the United Arab Emirates. Desalination 133:135–147
68. Glueckstern P, Priel M (1997) Optimized brackish water desalination plants with minimum impact on the environment. Desalination 108:19–26
69. Morton AJ, Callister IK, Wade NM (1997) Environmental impacts of seawater distillation and reverse osmosis processes. Desalination 108:1–10
70. Global Water Intelligence (2008) IDA desalination yearbook 2007–2008. Media Analytics, Oxford, UK
71. Mickols WE, Busch M, Maeda Y, Tonner J (2005) A novel design approach for seawater plants. IDA World Congress, Singapore
72. Mulder M (1996) Basic principles of membrane technology, 2nd edn. Kluwer, Dordrecht/Boston
73. Tufenki N, Elimelech M (2004) Correlation equation for predicting single-collector efficiency in physicochemical filtration in saturated porous media. Environ Sci Technol 38:529–536
74. Yao KM, Habibian MT, O'Melia CR (1971) Water and waste water filtration: concepts and applications. Environ Sci Technol 5:1105–1112
75. Elimelech M, Gregory J, Jia X, Williams RA (1995) Particle deposition and aggregation: measurement, modelling, and simulation. Butterworth-Heinemann, Oxford
76. Rajagopalan R, Tien C (1976) Trajectory analysis of deep-bed filtration with sphere-in-cell porous-media model. AIChE Journal 22:523–533
77. Shin JY, O'Mella CR (2006) Pretreatment chemistry for dual media filtration: model simulations and experimental studies. Water Sci Technol 53:167–175

78. Soltanieh M, Gill WN (1981) Review of reverse osmosis membranes and transport models. Chem Eng Commun 12:279–363
79. Chen KL, Song L, Ong SL, Ng WJ (2004) The development of membrane fouling in full-scale RO processes. J Membr Sci 232:63–72
80. Visvanathan C, Ben Aim R (1989) Studies on colloidal membrane fouling mechanisms in crossflow microfiltration. J Membr Sci 45:3–15
81. Kim YM, Lee YS, Lee YG, Kim SJ, Yang DR, Kim IS, Kim JH (2009) Development of a package model for process simulation and cost estimation of seawater reverse osmosis desalination plant. Desalination 247:326–335
82. Kim SJ, Lee YG, Cho KH, Kim YM, Choi S, Kim IS, Yang DR, Kim JH (2009) Site-specific raw seawater quality impact study on SWRO process for optimizing operation of the pressurized step. Desalination 238:140–157
83. Kim SJ, Lee YG, Oh S, Lee YS, Kim YM, Jeon MG, Lee S, Kim IS, Kim JH (2009) Energy saving methodology for the SWRO desalination process: control of operating temperature and pressure. Desalination 247:260–270
84. Lee YG, Lee YS, Jeon JJ, Lee S, Yang DR, Kim IS, Kim JH (2009) Artificial neural network model for optimizing operation of a seawater reverse osmosis desalination plant. Desalination 247:180–189
85. Kim SJ, Oh S, Lee YG, Jeon MG, Kim IS, Kim JH (2009) A control methodology for the feed water temperature to optimize SWRO desalination process using genetic programming. Desalination 247:190–199
86. Lee YG, Kim DY, Kim YC, Lee YS, Jung DH, Park M, Park SJ, Lee S, Yang DR, Kim JH (2010) A rapid performance diagnosis of seawater reverse osmosis membranes: simulation approach. Desalin Water Treat 15:11–19
87. El-Halwagi MM (1992) Synthesis of reverse-osmosis networks for waste reduction. AIChE Journal 38:1185–1198
88. Maskan F, Wiley DE, Johnston LPM, Clements DJ (2000) Optimal design of reverse osmosis module networks. AIChE Journal 46:946–954
89. Nemeth JE (1998) Innovative system designs to optimize performance of ultra-low pressure reverse osmosis membranes. Desalination 118:63–71
90. Voros N, Maroulis ZB, Marinos-Kouris D (1996) Optimization of reverse osmosis networks for seawater desalination. Comput Chem Eng 20:S345–S350
91. Zhu M, El-Halwagi MM, Al-Ahmad M (1997) Optimal design and scheduling of flexible reverse osmosis networks. J Membr Sci 129:161–174
92. See HJ, Vassiliadis VS, Wilson DI (1999) Optimisation of membrane regeneration scheduling in reverse osmosis networks for seawater desalination. Desalination 125:37–54
93. Voros NG, Maroulis ZB, Marinos-Kouris D (1997) Short-cut structural design of reverse osmosis desalination plants. J Membr Sci 127:47–68
94. Lu Y-Y, Hu Y-D, Xu D-M, Wu L-Y (2006) Optimum design of reverse osmosis seawater desalination system considering membrane cleaning and replacing. J Membr Sci 282:7–13
95. Lu Y-Y, Hu Y-D, Zhang X-L, Wu L-Y, Liu Q-Z (2007) Optimum design of reverse osmosis system under different feed concentration and product specification. J Membr Sci 287: 219–229
96. Lonsdale HK, Merten U, Riley RL (1965) Transport properties of cellulose acetate osmotic membranes. J Appl Polym Sci 9:1341–1362
97. Hatfield GB, Graves GW (1970) Optimization of a reverse osmosis system using nonlinear programming. Desalination 7:147–177
98. Tweddle TA, Thayer WL, Matsuura T, Fu-Hung H, Sourirajan S (1980) Specification of commercial reverse osmosis modules and predictability of their performance for water treatment applications. Desalination 32:181–198
99. Sirkar KK, Dang PT, Rao GH (1982) Approximate design equations for reverse osmosis desalination by spiral-wound modules. Ind Eng Chem Process Des Dev 21:517–527
100. Van Dijk JC, De Moel PJ, Van Den Berkmortel HA (1984) Optimizing design and cost of seawater reverse osmosis systems. Desalination 52:57–73
101. Evangelista F (1985) A short cut method for the design of reverse osmosis desalination plants. Ind Eng Chem Proc Des Dev 24:211–223
102. Sekino M (1993) Precise analytical model of hollow fiber reverse osmosis modules. J Membr Sci 85:241–252
103. Robertson MW, Watters JC, Desphande PB, Assef JZ, Alatiqi IM (1996) Model based control for reverse osmosis desalination processes. Desalination 104:59–68
104. Sekino M (1995) Study of an analytical model for hollow fiber reverse osmosis module systems. Desalination 100:85–97
105. Van der Meer WGJ, Riemersma M, Van Dijk JC (1998) Only two membrane modules per pressure vessel? Hydraulic optimization of spiral-wound membrane filtration plants. Desalination 119:57–64
106. Al-Bastaki NM, Abbas A (1999) Modeling an industrial reverse osmosis unit. Desalination 126:33–39
107. Al-Bastaki NM, Abbas A (2000) Predicting the performance of RO membranes. Desalination 132:181–187
108. Wilf M, Klinko K (2001) Optimization of seawater RO systems design. Desalination 138:299–306
109. Villafafila A, Mujtaba IM (2003) Fresh water by reverse osmosis based desalination: simulation and optimisation. Desalination 155:1–13
110. Helal AM, El-Nashar AM, Al-Katheeri E, Al-Malek S (2003) Optimal design of hybrid RO/MSF desalination plants Part I: modeling and algorithms. Desalination 154:43–66
111. Chatterjee A, Ahluwalia A, Senthilmurugan S, Gupta SK (2004) Modeling of a radial flow hollow fiber module and estimation of model parameters using numerical techniques. J Membr Sci 236:1–16
112. Marcovecchio MG, Aguirre PA, Scenna NJ (2005) Global optimal design of reverse osmosis networks for seawater desalination: modeling and algorithm. Desalination 184:259–271
113. Abbas A, Al-Bastaki N (2005) Modeling of an RO water desalination unit using neural networks. Chem Eng J 114:139–143

114. Geraldes V, Pereira NE, Norberta de Pinho M (2005) Simulation and optimization of medium-sized seawater reverse osmosis processes with spiral-wound modules. Ind Eng Chem Res 44:1897–1905
115. Senthilmurugan S, Gupta SK (2006) Modeling of a radial flow hollow fiber module and estimation of model parameters for aqueous multi-component mixture using numerical techniques. J Membr Sci 279:466–478
116. Moch & Assiciates B.R.E. (2002) WTCcost, water treatment cost estimation program, US Bureau of Reclamation
117. Farrell MD, Cimino J (2003) Reverse osmosis desalination cost planning model. In: Proceedings of IDA World congress on desalination and water reuse, Bahamas
118. Van der Bruggen B, Vandecasteele C (2002) Distillation vs. membrane filtration: overview of process evolutions in seawater desalination. Desalination 143:207–218
119. Abdella DL (1994) Reverse osmosis desalination. Marine Technol 31:195–200
120. McGinnis RL, Elimelech M (2008) Global challenges in energy and water supply: the promise of engineered osmosis. Environ Sci Technol 42:8625–8629
121. Cath TY, Childress AE, Elimelech M (2006) Forward osmosis: principles, applications, and recent developments. J Membr Sci 281:70–87
122. Miller JE, Evans LR (2006) Forward osmosis: a new approach to water purification and desalination. Sandia National Laboratories, Albuquerque. Livermore
123. McCutcheon JR, Elimelech M (2006) Influence of concentrative and dilutive internal concentration polarization on flux behavior in forward osmosis. J Membr Sci 284: 237–247
124. McCutcheon JR, McGinnis RL, Elimelech M (2005) A novel ammonia-carbon dioxide forward (direct) osmosis desalination process. Desalination 174:1–11
125. Chung TS, Zhang S, Wang KY, Su J, Ling MM (2011) Forward osmosis processes: yesterday, today and tomorrow. Desalination (doi:10.1016/j.desal.2010.12.019)
126. Achilli A, Childress AE (2010) Pressure retarded osmosis: from the vision of Sidney Loeb to the first prototype installation – review. Desalination 261:205–211
127. Phillip WA, Yong JS, Elimelech M (2010) Reverse draw solute permeation in forward osmosis: modeling and experiments. Environ Sci Technol 44:5170–5176
128. Lee KL, Baker RW, Lonsdale HK (1981) Membranes for power generation by pressure-retarded osmosis. J Membr Sci 8: 141–171
129. Jung DH, Lee J, Kim DY, Lee YG, Park M, Lee S, Yang DR, Kim JH (2011) Simulation of forward osmosis membrane process: effect of membrane orientation and flow direction of feed and draw solutions. Desalination 277:83–91
130. Tiraferri A, Yip NY, Phillip WA, Schiffman JD, Elimelech M (2011) Relating performance of thin-film composite forward osmosis membranes to support layer formation and structure. J Membr Sci 367:340–352
131. Park M, Lee JJ, Lee S, Kim JH (2011) Determination of a constant membrane structure parameter in forward osmosis processes. J Membr Sci 375:241–248
132. Mallevialle JL, Odendaal PE, Wiesner MR (1996) Water treatment membrane processes. McGraw-Hill, New York/London
133. Mi B, Elimelech M (2010) Organic fouling of forward osmosis membranes: fouling reversibility and cleaning without chemical reagents. J Membr Sci 348:337–345
134. Elimelech M (2007) Yale constructs forward osmosis desalination pilot plant. Membrane Technol 2007:7–8
135. Tan CH, Ng HY (2008) Modified models to predict flux behavior in forward osmosis in consideration of external and internal concentration polarizations. J Membr Sci 324:209–219
136. Phuntsho S, Shon HK, Hong S, Lee S, Vigneswaran S (2011) A novel low energy fertilizer driven forward osmosis desalination for direct fertigation: evaluating the performance of fertilizer draw solutions. J Membr Sci 375:172–181
137. Bodell BR (1963) Silicone rubber vapor diffusion in saline water distillation. US Patent 285032
138. Findley ME (1967) Vaporization through porous membranes. IandEC Process Des Dev 6:226–230
139. Gore DW (1982) Gore-tex membrane distillation. In: Proceedings of the tenth annual convention of the water supply improvement association, Honolulu
140. Andersson SI, Kjellander N, Rodesjo B (1985) Design and field tests of a new membrane distillation desalination process. Desalination 56:345–354
141. El-Bourawi MS, Ding Z, Ma R, Khayet M (2006) A framework for better understanding membrane distillation separation process. J Membr Sci 285:4–29
142. Schneider K, Van Gassel TJ (1984) Membrane distillation. Membrane distillation 56:514–521
143. Lawson KW, Lloyd DR (1997) Membrane distillation. J Membr Sci 124:1–25
144. Banat FA, Simandl J (1994) Theoretical and experimental study in membrane distillation. Desalination 95:39–52
145. Bandini S, Gostoli C, Sarti GC (1992) Separation efficiency in vacuum membrane distillation. J Membr Sci 73:217–229
146. Khayet M, Godino P, Mengual JI (2000) Theory and experiments on sweeping gas membrane distillation. J Membr Sci 165:261–272
147. Kimura S, Nakao SI, Shimatani SI (1987) Transport phenomena in membrane distillation. J Membr Sci 33:285–298
148. Martínez-Díez L, Vázquez-González MI (1999) Temperature and concentration polarization in membrane distillation of aqueous salt solutions. J Membr Sci 156:265–273
149. Schofield RW, Fane AG, Fell CJD (1987) Heat and mass transfer in membrane distillation. J Membr Sci 33:299–313
150. Zhu C, Liu G (2000) Modeling of ultrasonic enhancement on membrane distillation. J Membr Sci 176:31–41
151. Khayet M (2011) Membranes and theoretical modeling of membrane distillation: a review. Adv Colloid Interface Sci 164:56–88
152. Hogan PA, Sudjito A, Fane AG, Morrison GL (1991) Desalination by solar heated membrane distillation. Desalination 81:81–90

153. Weyl PK (1967) Recovery of demineralized water from saline waters, United States Patent 3340186
154. Kim S, Cho D, Lee MS, Oh BS, Kim JH, Kim IS (2009) SEAHERO R&D program and key strategies for the scale-up of a seawater reverse osmosis (SWRO) system. Desalination 238:1–9
155. Park M, Lee YS, Lee YG, Kim JH (2010) SeaHERO core technology and its research scope for a seawater reverse osmosis desalination system. Desalin Water Treat 15:1–4
156. Criscuoli A, Drioli E (1999) Energetic and exergetic analysis of an integrated membrane desalination system. Desalination 124:243–249
157. Phattaranawik J, Jiraratananon R, Fane AG (2003) Heat transport and membrane distillation coefficients in direct contact membrane distillation. J Membr Sci 212:177–193

Desertification and Impact on Human Systems

David Mouat, Judith Lancaster
Division of Earth and Ecosystem Sciences,
Desert Research Institute, Reno, NV, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Desertification and Human Systems
Discussion
Future Directions
Bibliography

Glossary

Adaptation changes made by organisms, including humans, to enable them to be more suitable for different conditions or situations.

Desertification the gradual change of habitable land which affects soils, flora, and fauna, and reduces productivity and an ecosystem's ability to adapt.

Drought a condition under which human demand for water exceeds its natural variability, usually considered to result from an extended period of reduced rainfall.

Drylands include arid, semi-arid, and dry sub-humid areas, and have a ratio of mean annual precipitation to mean annual potential evapotranspiration ranging between 0.05 and 0.65.

Ecosystem goods and services the benefits people obtain from ecosystems. These may be divided into provisioning (food and water), regulating (controlling floods and diseases), cultural (recreational, spiritual), and supporting services such as nutrient cycling.

Land degradation a temporary or permanent lowering of the biological and/or economic productive capacity of land, including changes to soil and vegetation.

Resilience the ability to recover quickly from change. Returning to the original state is often mentioned as part of a definition, but there are aspects of adaptation linked to resilience which implies that some change of state may be part of the process.

Vulnerability a condition of being at risk, and some difficulty in dealing with the situation.

Definition of the Subject

Desertification is not a new phenomenon, but it is seen to be increasing in severity and extent, with an estimated 10–20% of drylands degraded in 2005 [1]. The term "desertification" originates from the 1940s [2] although it was not widely recognized until the West African sub Sahelian droughts of the 1960s and 1970s. As Dietz and coauthors [3] discuss, the West African area has recovered from those droughts, not only because natural fluctuations in climate resulted in more rainfall but, and most importantly, because of human adaptation to the desertification "situation." In fact, desertification is not caused by a lack of rain – but by numerous interrelated contributing factors from both natural and social systems, which are acknowledged in this definition:

"Land degradation in arid, semi-arid and dry sub-humid areas resulting mainly from negative human impacts combined with difficult climatic and environmental conditions" [4].

Introduction

Faced with intensifying land degradation at the global scale, the United Nations Convention to Combat Desertification (UNCCD) was adopted in 1994, emphasizing sustainable development at the community level [5]. Regional and national assessments have been carried out [6] and research addresses many

issues. The UNCCD is committed to "bottom-up" action, recognition of traditional knowledge, and the importance of women, yet gaps among scientists, policy- and decision-makers, and communities still exist in many dryland areas hampering efforts to address this environmental, economic, and social problem.

Desertification affects arid, semi-arid, and dry subhumid regions, occurs on every continent, and affects more than two billion people who live in these areas. These are inherently fragile ecosystems often close to the tipping point between continued production of ecosystem goods and services [1] and a spiral of change leading to barren landscapes where people can no longer survive. Dryland regions tend to receive minimal attention from local or national agencies, and are typically poorly provided for in terms of education, health, and infrastructure services. Dryland peoples are thus marginalized [7], going unnoticed unless war, natural disasters, or famine draws attention to them. Poverty and desertification are strongly linked [1], and this situation is exacerbated by the political instability of many dryland regions.

Desertification and Human Systems

Changes in Production Systems

The extent, severity, and impacts of desertification vary in both space and time, driven by pressures people put on dryland ecosystems combined with the intensity of aridity [1]. Therefore, dryland productivity , the provision of goods and services – such as water, wood for fuel and building, and fodder for grazing – also varies. People have developed a range of coping mechanisms in response to the natural fluctuations of ecosystems, which include nomadism, shifting cultivation, and surplus accumulation. As pointed out by Reynolds and Stafford Smith [8] dryland peoples are not the "problem" in desertification, nor are they "victims," they are one part of an integrated system, and their responses to environmental change vary depending upon the severity, duration, and scale of the change. However, growth in both population and poverty may render previously effective coping mechanisms inadequate, and result in increased vulnerability to hunger, disease, and political pressures. With two billion people involved, these issues are serious and are likely to become more so as the implications of climate change are factored in.

Long-term trends to more intense and longer droughts and overall drying have been observed in the Sahel, Mediterranean, southern Africa, and areas in southern Asia [9] over the period 1900–2005. Climate change models for Africa indicate that temperatures are likely to increase 0.2–0.5°C per decade [10], which will contribute to increasing rainfall variability by more than one standard deviation from normal in many areas [11], and will, in turn, significantly decrease perennial surface runoff. In the Western Cape of South Africa, for example, up to half the present perennial water supply is likely to be lost – even based on a relatively optimistic climate model used by de Wit and Stankiewicz [10]. The Intergovernmental Panel on Climate Change (IPCC) [11] found that decreases in runoff totaled about 17% during the 1990s, which indicates that the trend has already started.

The IPCC Working Group 2 [in 9] reports that reduced rainfall when coupled with temperature increases, not only may lead to diminished surface water availability and less recharge, but also affects plants – to the point where current crop varieties may suffer reduced yields or not produce at all. IPCC Working Group 1 suggest that reduced rainfall in some areas juxtaposed with increased flooding in others, may make rain fed agriculture, which is practiced in many dryland areas, a precarious undertaking [9]. Rangeland, or natural pastures, upon which pastoralists depend for stock grazing are also likely to be affected with reductions in forage quantity and extent, and water points will become less reliable and productive.

These changes to natural resource viability will have, or are having, impacts beyond food shortages for dryland peoples. Competition for resources may weaken reciprocal arrangements which are inherent parts of coping mechanisms – increasing vulnerability, the possibility of conflict, migration, and the failure of social institutions such as tenure and inheritance systems, markets, and subsidies [8].

Response Strategies

Response strategies and social resilience to environmental change have been shown to be partially dependent upon the range of response options available – often at the individual household level [12]. In their analysis of building social resilience in arid ecosystems, Vogel and

Smith [12] suggest that households in the most marginal of environments tend to be poor in terms of options, lack resources for making changes, and therefore have not developed a diversified suite of production activities. Some level of household stratification exists in many parts of the southern African examples discussed by Vogel and Smith [12] and other authors in this entry.

In Namibia, unpredictable precipitation and concomitant productivity can result in cycles of abundance and paucity of natural resources and dryland agriculture crops. In the good years everyone "banks" for the future, whether it is literally in monetary terms as is the case for the wealthier individuals, or in terms of famine food stored for lean times by poorer families and communities. The stored food option is adequate until the natural cycles are changed by a prolonged drought (or climate change perhaps) when options run out, and vulnerability sets in.

Croppers and herders in Senegal have different levels of social resilience and decision-making options, yet coexist and have developed some intergroup adaptive strategies [13]. In describing their Social Resilience Model, Bradley and Grainger [13] discuss both social and environmental resilience, concluding that desertification in their study region was less severe than might be expected due to human behavior [cf.3]. By implication, adaptive strategies in the region are successful and maintain livelihoods despite climatic fluctuations, impacts of decisions by policy-makers, and the inherent variability of ecosystem resilience in time and space.

Also in Senegal, the National Action Plan initiated as part of the country's ratification of the UNCCD, established a national forum on the involvement of women. In Kenya, 30–50% of the participants in the formulation of the national action plan were women, and northeast Brazil is particularly active in promoting women's role in sustainable development [14]. Addressing gender inequalities in land tenure, inheritance, and decision making and recognizing the importance of women for dryland communities and ecosystems is a positive trend, which is gathering momentum in many countries.

Adaptation Systems and Improving Livelihoods

Most adaptation systems are of grass roots origin and local implementation, but this does not preclude their inclusion or performance within more formal organizational structures. National, Regional, and Subregional Action Programmes are key to implementing the United Nations Convention to Combat Desertification and although such programs might appear to be "imposed" or "top down," the UNCCD has recognized the importance of community level involvement, local knowledge, the role of women, and the value of synergistic activities between other UN programs [15]. The programs implemented by the UNCCD, and many in-country NGOs, operate at local level and aim to improve livelihoods within the context of desertification, rather than (or sometimes as well as) instigating remediation.

In several areas of northwestern China, desertification has been reduced over the past 5 years due to the application of new policies designed to change human activity and protect the environment. The "Grain for Green" and "Grazing Prohibition" policies are improving the environment, but other challenges remain – such as how to improve farmer's income and develop the economy. These questions were driving forces behind Zhou and Mouat's hypothetical study [16], which identified five uncertainties facing the Minqin region:

- Will there be sufficient water for agricultural and domestic use?
- What effect will climate change have?
- Is soil salinization likely to increase?
- What changes to the economy will occur, possibly as a result of government policy?
- Are land use patterns going to change?

Based upon the present land use status, one of the potential alternative futures for the region developed by Zhou and Mouat [16] focuses upon the development of high tech industry in the city of Wuwei. As a result of this, the city expands somewhat into the agricultural area, but the rural population remains on the land while taking advantage of new opportunities in the city. A policy to import high water use commodities initiated by the administration in Wuwei, combined with mandated water conservation, results in lowered water use for that area – and agriculture downstream, around Minqin, improves. Although based upon a hypothetical study, such diversification is an example of successful social resilience, which, in

conjunction with policy decisions, also enhances environmental resilience and minimizes increased desertification [cf. 13].

Management strategies – for land, water, and human resources – are critical for human existence in all ecosystems, but particularly those such as drylands which are fragile and where human–environmental interactions may have considerable implications. Traditional policies and organizations for water resource management are still effective for managing wells in southern Ethiopia [17]. However, Emiru [17] reports that newly constructed water points are often administered poorly due to newly instigated management committees. Not all "new" management strategies cause problems, as evidenced by the success of Conservancies (community-based conservation efforts, frequently targeting natural resources) [e.g., 18]. Many countries in Africa are adopting the Conservancy approach, which is usually focused on promoting conservation of natural resources and wildlife, plus sustainable utilization and sharing of resources. These organizations are a diversification from the single-family approach to management, and, recalling the arguments put forward by Bradley and Grainger [13], it is those who are able to maintain options and diversify who are most successful in living with desertification.

In western Senegal there are several strategies which are adopted by the pastoralist and agricultural groups in times of reduced productivity [13] – what is particularly interesting is how the groups vary in their perceptions and adaptations. The pastoralists move between their grazing, watering, and trade sites on a seasonal and long distance basis, both as an anticipatory and a response strategy. During these migrations, social networks and use rights are maintained on a regular basis. Both agriculturalists and pastoralists accumulate surpluses. Agriculturalists expand the area under cultivation, pastoralists build up herds with the idea that at least some of the animals would survive a future drought.

Both groups have some diversity in livelihoods including collection of forest products, trade, and spiritualist activities in addition to cropping and/or raising livestock. However, Bradley and Grainger [13] report that the pastoralists diversified more easily and had a confident approach to this adaptive strategy. In response to changing environmental conditions both groups maintained their customary production mode as long as possible, but, under pressure, pastoralists would increase their mobility, diversify stock type, or enter the gum-arabic trade. However, the response by agriculturalists is typically for some of the group to migrate to urban areas to look for work. It would appear that the pastoralists adapt readily and are innovative in their solutions, possibly as a result of the mobility and therefore flexibility inherent in their "normal" lifestyle.

Discussion

Human adaptations which improve both social and environmental resilience – such as diversification and development of reciprocal networks – are likely to be more successful, for human and ecosystems, than abandoning the land and migrating to cities. This latter strategy increases the potential for desertification, adds to the problem of urban poor, and presents a security risk for many regions.

Challenges from climate change, decreased productivity and biological diversity, poverty, and changing social institutions face the populations of the world's drylands, but Reynolds et al. [7] suggest a framework which can be used by managers and policy-makers to address the complexity of interlinked systems, and recognize what is important to change and where research can help. Based on the major characteristics of dryland ecosystems and social systems, Reynolds et al. [7] show how these relate to factors such as the maintenance of local environmental knowledge as key to functional coadaptation of both systems. This relationship is extrapolated to implications for research, management, and policy – for the local knowledge example above, this focuses on accelerating integration of science and local environmental knowledge at local and regional levels for management and policy.

Global climate change models operate at scales that do not permit a local level of analysis, and given the great spatial variability of dryland ecosystems there will be some areas which are less affected than others – even within the same ecosystem and climate regime [19]. This internal variability will impose additional social pressures upon dryland peoples, adding a further element of uncertainty to society, sustainability, and

human and environmental security. It is imperative that desertification is thought of as a process, a problem, a phenomenon that affects people not just in those places directly involved but globally as well.

Future Directions

In this entry, the process of desertification, especially as it integrates biophysical system and human use systems has been described. As Reynolds and Stafford Smith have stated, people are neither the problem nor the victims, they are a part of an integrated system. Causes and effects have been discussed as have International efforts to "combat," mitigate, and prevent desertification (UNCCD). Finally, response strategies and adaptation systems are discussed.

What will be the complexion of efforts to deal with desertification as the twenty-first century progresses? Increasing populations and concomitant competition for resources including water and arable land will require strategies to ameliorate a likely worsening situation. These include understanding how to use water and other resources more efficiently, developing salt and drought resistant crops, and developing local and green energy resources. As pointed out, at the local level, people faced with increasing land degradation and decreasing livelihoods adapt, revolt, migrate, or die. International, national, and NGOs are increasingly striving to communicate success stories across communities. These same organizations, at the very least, are drawing awareness to the problems that people living in drylands face. They are also raising awareness of the importance of developing bottom-up approaches, involving women and of using traditional knowledge in those same efforts. Developing infrastructure will assist. While early warning systems for forecasting drought have been around for decades, similar early warning systems for forecasting degradation are only in a nascent state.

Increasing our understanding of the causes and processes of desertification is important but so is the importance of figuring out how to use what is already known. It is important to ask questions about how drylands might change in the future including plausible alternatives. Assessing and analyzing the alternatives will allow one to pick pathways of development that will allow for positive futures. In our view, it is important to remain focused on the sustainability of desert ecosystems and for the people who live there, and it is essential to focus on the impacts of climate change on human use systems in the context of sustainability. While much attention is focused on ameliorating and mitigating the effects of desertification, it is important to prevent drylands from becoming desertified in the first place.

Bibliography

1. MEA (Millennium Ecosystem Assessment) (2005) Ecosystems and human well-being: desertification synthesis. World Resources Institute, Washington, p 25
2. Aubreville A (1949) Climats, forest, et desertification de l'Afrique Tropicale. Societe de Editions Geographiques, Maritime et Coloniales, Paris, p 255
3. Dietz AJ, Ruben R, Verhagen A (Eds) (2004) The impact of climate change on drylands: with a focus on West Africa. Environment and Policy Series, vol 39. Springer, New York, p 468
4. UNCED (United Nations Conference on Environment and Development) (1992) Earth summit agenda 21: programme of action for sustainable development. United Nations Department of Public Information, New York
5. The Convention. http://www.unccd.int/convention/menu.php. Accessed Nov 2010
6. Middleton N, Thomas D (eds) (1997) World atlas of desertification, 2nd edn. Arnold, London, p 182
7. Reynolds JF, Stafford Smith DM, Lambin EF, Turner BL II, Mortimore MN, Batterbury SPJ, Downing TE, Dowlatabadi H, Fernandez RJ, Herrick JE, Huber-Sannwald E, Jiang H, Leemans R, Lynam T, Maestre FT, Ayarza M, Walker B (2007) Global desertification: building a science for dryland development. Science 316:847–851, and www.sciencemag.org. May 11, 2007
8. Reynolds JF, Stafford Smith DM (2002) Do humans cause deserts?, pp 1–21. In: Reynolds JF, Stafford Smith DM (eds) Global desertification: do humans cause deserts?. Dahlem University Press, Berlin, p 437
9. CCDC (Commission on Climate Change and Development) (2008) Climate change and drylands. http://www.ccdcomission.org. Accessed Nov 2010
10. de Wit M, Stankiewicz J (2006) Changes in surface water supply across Africa with predicted climate change. Science 311:1917–1921
11. IPCC (Intergovernmental Panel on Climate Change) (2001) - Climate change: impacts, adaptation and vulnerability. World Meteorological Organization, Geneva
12. Vogel CH, Smith J (2002) Building social resilience in arid ecosystems, pp 149–165. In: Reynolds JF, Stafford Smith DM (eds) Global desertification: do humans cause deserts?. Dahlem University Press, Berlin, p 437

13. Bradley D, Grainger A (2004) Social resilience as a controlling influence on desertification in Senegal. Land Degrad Dev 15:451–470
14. (2004) Women and desertification: a dynamic relationship. Chapter 4, in UNEP, 2004, Women and the Environment, pp. 49–59. http://www.unep.org/pdf/women/chapterfour.pdf
15. UNCCD. http://www.unccd.int/main.php. Accessed Nov 2010
16. Zhou L, Mouat D (2007) Whose decisions impact land use change: the people or the government? Thoughts from Northwest China. Unpublished material
17. Emiru N (2010) Role of traditional institutions in water resource governance in the Borana lowlands, southern Ethiopia. Haramata 55:13–15
18. Who is CANAM. http://www.canam.iway.na/home.htm. Accessed Nov 2010
19. IUCC (Information Unit on Climate Change) (1993) Desertification and climate change. UNEP, Châtelaine

Direct Hydrocarbon Solid Oxide Fuel Cells

Michael van den Bossche[1], Steven McIntosh[1,2]
[1]Department of Chemical Engineering, University of Virginia, Charlottesville, VA, USA
[2]Department of Chemical Engineering, Lehigh University, Bethlehem, PA, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Background
Copper-Ceria
Lanthanum Chromates
Double Perovskites
Strontium Titanates
Other Materials of Interest
Future Directions
Bibliography

Glossary

Electrical conductivity Also referred to as total conductivity. This total conductivity has three components: electronic n-type (electron charge carriers), electronic p-type (electron hole charge carriers), and ionic conductivity.

Electrocatalytic reaction Catalyzed reaction that involves charge transfer.

Humidified fuel In this text, humidified fuel refers to fuel that has been passed through room temperature water and hence contains 3 mol% H_2O.

OCP Open Circuit Potential. The potential difference measured between anode and cathode for an SOFC with an open electronic circuit. If the redox reactions occurring at the electrodes are known, the OCP can be predicted using the Nernst equation.

Overpotential The decrease in the maximum driving force (OCP), in a working SOFC. This is typically caused by charge transfer processes in the electrodes.

Oxygen stoichiometry The oxygen content of a solid oxide material. Most of the materials of interest in this study exhibit variable oxygen stoichiometry as a function of oxygen partial pressure and temperature while maintaining the cation structure.

Polarization resistance The electrical resistance of the electrode of an electrochemical device upon polarization.

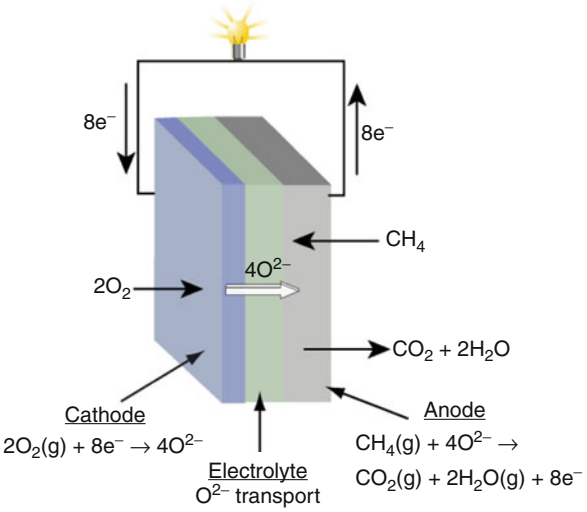
SOFC Solid Oxide Fuel Cell. A fuel cell characterized as having a solid oxide electrolyte that separates the air electrode (cathode) from the fuel electrode (anode).

TEC Thermal Expansion Coefficient. Sometimes referred to as the coefficient of thermal expansion (CTE) and refers to the total expansion of the lattice cell as a function of temperature. For oxides with variable oxygen stoichiometry, the TEC consists of the thermal expansion and the chemical expansion. The latter is the expansion due to changes in oxygen stoichiometry. In this text it refers to the total expansion.

Definition of the Subject

Solid Oxide Fuel Cells (SOFCs) are one of the most promising technologies for future efficient conversion of the chemical energy stored in fuels to electrical energy. One of the primary advantages of SOFCs is the potential to operate with a wide variety of fuels. While H_2 is the fuel of choice for most fuel cells, operation with fossil-derived and bio-derived hydrocarbon fuels would bypass the costly requirement of a new H_2 infrastructure, and accelerate adoption of fuel

cell technology. This is feasible as SOFCs transport oxygen anions from the air electrode (cathode) to the fuel electrode (anode). The primary barrier to the realization of fuel flexible SOFCs is the anode material set. Traditional SOFC anodes are based on Ni composites. While these are very efficient for H_2 and CO fuels, Ni catalyzes graphite formation from dry hydrocarbons, leading to rapid degradation of cell performance and possible mechanical failure. This motivates the development of new anode materials and composites. The principle requirements of an anode are oxygen anion conductivity, electronic conductivity, and electrocatalytic activity toward the desired reaction. This entry reviews the primary issues in direct hydrocarbon anode development and discusses the new materials and composites that have been developed to meet this challenge.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 1 Solid Oxide Fuel Cell (SOFC) operating principle

Introduction

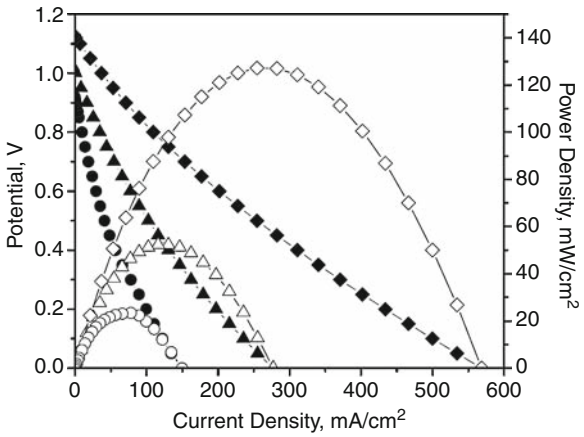
Traditional combustion-based methods of electrical power generation convert the chemical energy of a fuel to electrical energy via an intermediate thermal step. In contrast, electrochemical systems, such as batteries and fuel cells, promise increased efficiency through the direct conversion of chemical to electrical energy. This serves to reduce fuel demand and decrease associated greenhouse gas emissions per unit of electrical energy output. The defining difference between a fuel cell and a battery is that a battery is a closed system with a finite amount of "fuel," whereas a fuel cell is an open flow system with continuous feed of fuel and oxidant. Where a battery becomes depleted and must be recharged, a fuel cell continues to generate electrical power as long as both fuel and oxidant are supplied to the cell. Therefore, fuel cells are an efficient alternative for continuous power generation from a chemical fuel source.

There are many types of fuel cells with the primary differentiators between each type being the ion transported across the electrolyte and the operating temperature range. This entry focuses on solid oxide fuel cells (SOFCs) [1–5]. Like all fuel cells, SOFCs consist of three main components: a cathode (air electrode), an anode (fuel electrode), and an electrolyte (purely ion conducting membrane between the cathode and anode). Figure 1 is a schematic of the SOFC operating principle, showing operation with CH_4 fuel. The distinguishing feature of SOFCs is that the electrolyte is a dense, oxygen anion conducting but electronically insulating solid oxide. The SOFC operating mechanism is based upon the transport of oxygen anions from an oxygen-rich environment at the cathode (typically air) to an oxygen-lean fuel environment at the anode. The fuel and air chambers are physically and electronically separated by the dense electrolyte and the entire cell is sealed to prevent gas-phase mixing of fuel and air. The anode and cathode are ionically connected via the electrolyte, and electrically connected through an external circuit.

Under operation, electrons from the external circuit are utilized in the electrocatalytic reduction of molecular oxygen to oxygen anions at the cathode. These anions migrate through the electrolyte from the cathode to the anode. Fuel is electrocatalytically oxidized at the anode (consuming the oxygen anions) and the liberated electrons flow back to the cathode via external circuit, providing useful work. The electrochemical driving force for this process is the oxygen chemical potential difference between the cathode and anode compartments. The maximum theoretical thermodynamic driving force, the theoretical open circuit potential (OCP), can be calculated from the Nernst equation. This driving force is consumed as current is drawn

from the cell. The current generated for a given potential drop from OCP, or overpotential, is determined by the magnitude of the resistive processes within the cell electrodes and electrolyte. Assuming a constant cell temperature, electrolyte material, and electrolyte thickness, the cell current is maximized by minimizing losses associated with the reaction and transport in the electrodes. Since power is the product of potential and current, maximizing current at a high potential leads to maximum power output. The typical performance plot for an operating SOFC is a potential-current, or voltage-current (V-I) plot with accompanying power-current curve. An example is shown in Fig. 2 for an SOFC operating with H_2, CH_4, and n-C_4H_{10} fuels at 700°C.

The most commonly utilized SOFC electrolyte is 8 mol% yttria-stabilized zirconia (YSZ). YSZ is utilized as it is an almost pure oxygen ion conductor with acceptable conductivity, and is stable in both anode and cathode gas environments. The cathode [7] is a mixture of an electron or mixed oxygen ion and electron conducting oxide and the YSZ electrolyte. Typical cathode materials include $La_{0.2}Sr_{0.8}MnO_{3\pm\delta}$ (LSM) and the $La_{1-x}Sr_xCo_{1-y}Fe_yO_{3-\delta}$ (LSCF) family. The traditional anode is a combination of Ni metal and YSZ, where Ni provides electronic conductivity and electrocatalytic activity [3].



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 2 Cell potential (*closed symbols*) and power density (*open symbols*) as a function of current density in dry H_2 (*diamonds*), C_4H_{10} (*triangles*), and CH_4 (*circles*) fuels at 700°C. The cell had an LSM/YSZ cathode, 50 μm thick YSZ electrolyte, and 8/35/57 wt% $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-\delta}$/Cu/YSZ anode. (Reproduced with permission from [6]. Copyright 2008, The Electrochemical Society)

Due to the high activation energy of ion transport in oxides, current SOFCs require operating temperatures greater than ~700°C. This high-temperature operation is advantageous in increasing efficiency, enabling the use of relatively inexpensive transition metal and metal oxide catalysts, and removing the issue of CO poisoning that plagues low temperature cells. In addition, the total system efficiency can be increased by utilizing the waste heat to generate more electrical power in a secondary turbine process. However, higher operating temperatures present challenges with regard to balance of plant material costs, startup and shutdown times, and system lifetime.

Full optimization of single cells and cell stacks requires optimization of all components, which is far beyond the scope of a single entry. There are a number of excellent reviews available in the literature that discuss SOFC development [2, 8–10]. The focus of our discussion is the direct utilization of hydrocarbon fuels in SOFC anodes. As such, the discussion here is limited to the anode material challenges and potential solution pathways specific to this goal.

The challenge for researchers is to design electrode materials and composites that provide sufficient ionic and electronic conductivity, and selectively catalyze hydrocarbon fuel oxidation. These materials must operate at high temperatures in highly reducing environments. This entry first provides background regarding the specific challenges and a brief introduction to the relevant materials chemistry. This is followed by more detailed discussions of possible solution pathways, grouped into broad material classes.

Background

Anodes for Hydrocarbon Fuels

Irrespective of the power generation route, any method that seeks to maximize electrical energy output via oxidation of a hydrocarbon fuel will emit the carbon content of the fuel as CO_2. This occurs irrespective of the number of intermediate chemical, electrochemical, or thermal steps. The goals for future power generation are thus to utilize both traditional fossil fuels and future

bio-derived hydrocarbons, minimize the amount of fuel used and CO_2 emitted per unit of electrical energy generated, and to generate the CO_2 in a high purity and more easily sequestered form.

Since SOFCs transport the oxidant (oxygen anions) to the fuel, Fig. 1, SOFCs can theoretically operate with any oxidizable fuel that can be supplied to the anode. The simplest and most commonly researched SOFC fuel is H_2 [2]. However, H_2 is not a primary energy source with ~96% of industrial hydrogen generated via energy intensive endothermic steam reforming of hydrocarbons, primarily CH_4. This is followed by high and low temperature water gas shift reactors to maximize H_2 production, and final product purification. The carbon content of the hydrocarbon is rejected as CO_2 at the point of H_2 generation. The remaining ~4% of industrial H_2 is generated by water electrolysis. The energy used to produce H_2, either by reforming or electrolysis, represents a parasitic load on the system that reduces efficiency and increases emissions. Furthermore, while hydrogen has a high mass-based energy density, the volumetric energy density is low and it is currently neither easily stored nor transported at the required scale. Direct utilization of hydrocarbon fuels would remove these barriers to fuel cell adoption and increase system efficiency.

Utilization of CO and H_2 mixtures can be considered a first step along the path to direct utilization of hydrocarbons. As with production of pure H_2, CO and H_2 mixtures are generated by hydrocarbon steam reforming. However, since the SOFC can utilize CO as a fuel, the water gas shift and H_2 purification steps are unnecessary. Hydrocarbon steam reforming can be performed either in an external reactor or internally on the SOFC anode itself. An external reactor operates as a typical hydrocarbon steam reforming system, while internal reforming requires addition of significant amounts of H_2O or O_2 to the fuel prior to feeding into the cell anode compartment. While both approaches remove the hydrogen production, storage, and transport demands, the system complexity increases and efficiency decreases compared to the promise of direct hydrocarbon utilization.

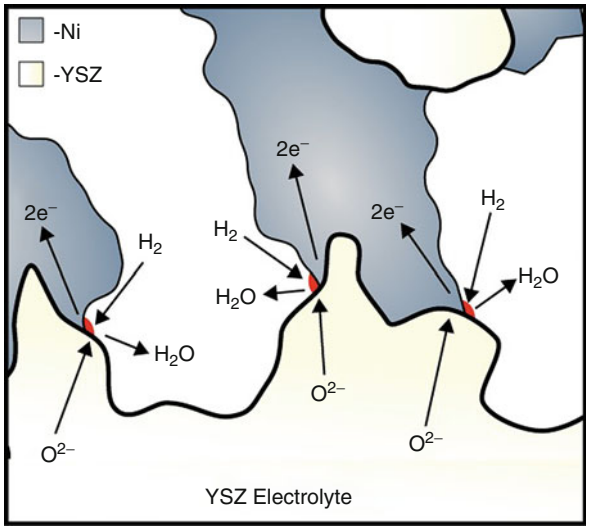
The goal for direct hydrocarbon anode design is to realize the complete electrochemical oxidation of the fuel to the total oxidation products H_2O and CO_2. This represents the full conversion of the chemical energy potential in the fuel to electrical energy in a single process unit. Where only one oxidation reaction is feasible for H_2, Eq. 1, hydrocarbons at high temperatures can undergo a large number of oxidation and polymerization reactions. The desired overall anode reaction is the total oxidation of fuel, shown in Eq. 2 for CH_4, as this represents complete utilization of the fuel and generates the maximum moles of electrons per mole of fuel consumed.

$$H_2 + O^{2-} \rightarrow H_2O + 2e^- \quad (1)$$

$$CH_4 + 4O^{2-} \rightarrow CO_2 + 2H_2O + 8e^- \quad (2)$$

By considering Eqs. 1 and 2, three clear requirements for SOFC anode materials can be derived. First, oxygen anion conductivity is required to transport the reactant oxygen anions to the reaction site. Second, the anode must selectively facilitate the desired electrocatalytic oxidation of the fuel. Third, facile electrical conductivity is required to transport the product electrons from the reaction site to the current collector wire. These material requirements are in addition to considerations of materials compatibility and stability during cell fabrication and operation, porosity in the electrode for fuel and product gas-phase diffusion, structural integrity, tolerance to impurities in the fuel and anode materials, and redox stability in case of accidental oxidation.

Current anodes for H_2 or humidified CO/H_2 fuelled SOFCs are Ni–YSZ composites. The electrolyte YSZ provides ionic conductivity, while metallic Ni is active for the oxidation of CO and H_2 and is a good electronic conductor. The operation of this anode is shown schematically in Fig. 3. Here operation with H_2 fuel is depicted, with the anode reaction given by Eq. 1. These electrodes are typically fabricated through slurry mixing and tape-casting of NiO and YSZ powders with or without pore former. The two materials are then co-fired at high temperature (>1,450°C) with the electrolyte. This is possible due to the high melting point of NiO (1,960°C) and lack of solid state reaction between NiO and YSZ. One limitation of utilizing two components is that the anode reaction can only take place at the interfaces between the Ni, YSZ, and gas phases as these are the only points at which all of the anode requirements are met – *the triple phase boundary (TPB)* highlighted red in Fig. 3. This disadvantage can



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 3 Schematic of Ni-YSZ based anode operating with H_2 fuel

be overcome by careful electrode design in terms of porosity and Ni/YSZ particle sizes and interconnectivity, and Ni-YSZ anodes do provide low overpotentials, and thus high cell power output, for H_2 and H_2/CO fuel mixtures.

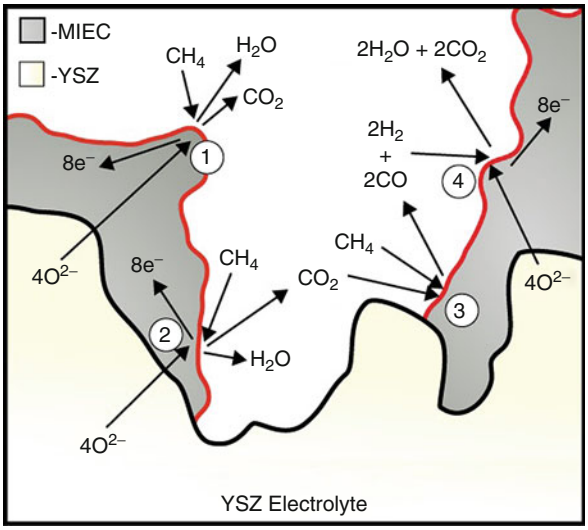
The primary barrier to direct hydrocarbon utilization with these traditional anode materials is the activity of Ni toward the formation of graphitic carbon from dry hydrocarbons [11]. These carbon deposits rapidly build up in the anode structure and can generate sufficient stress to fracture the cell. While carbon formation can be suppressed by careful selection of cell operating parameters [12–16], this requires operating an inherently unstable system. For example, Barnett and coworkers demonstrated CH_4 utilization on a Ni-YSZ composite anode by using the oxygen anion flux through the electrolyte to suppress carbon formation [14]. This is impractical for application as the cell must be stable under all potential operating conditions. Thus a considerable effort has been expended to design materials and composites to replace Ni as the electronic and electrocatalytically active component.

Oxidation of Hydrocarbons in SOFC

Almost all studies regarding direct hydrocarbon SOFCs show comparatively poor performance (lower OCP and higher polarization resistance) with hydrocarbon fuels when compared to H_2 fuel, Fig. 2. Since most of these tests are performed by switching fuel on the same cell, the drop in performance must be linked to the anode. It is possible that the increased polarization resistance may be due to lower diffusivity of the hydrocarbon fuels, but the electrodes are typically highly porous and the current density per unit area is relatively low. In addition, the oxidation of 1 mole of hydrocarbon fuel yields a significantly greater number of electrons than 1 mole of H_2 fuel (H_2, CH_4, and C_4H_{10} total oxidation yield 2, 8, and 26 moles of electrons, respectively). Furthermore, the cell OCP is an equilibrium, zero current, measurement and is therefore not directly influenced by gas diffusivity. Therefore, it is unlikely that gas diffusivity limits the performance for pure fuels at low conversion. The conclusion must then be that the anode electrocatalytic activity toward hydrocarbon oxidation is the primary factor in reduced SOFC performance.

An ideal anode material would possess sufficient ionic conductivity, electronic conductivity, and electrocatalytic activity. With all of these properties present in a single mixed ionic-electronic conducting (MIEC) phase, the entire anode surface within the active region close to the electrolyte would be active toward fuel oxidation. This is depicted for CH_4 fuel in Fig. 4 with the active surface highlighted in red. The thickness of this active region would be dictated by the relative rates of oxygen ion transport and surface reaction rate [17]. If the surface reaction selectivity were 100% toward total oxidation of hydrocarbon fuel, it may be anticipated that all reactions would follow pathway 1 in Fig. 4. In this case, oxygen anions migrate from the electrolyte to the MIEC phase, and are utilized to oxidize CH_4 to CO_2 and H_2O in an electrocatalytic surface reaction. The product gases diffuse out of the anode pores with product electrons conducted through the MIEC phase to the surface current collector.

While a number of groups are pushing toward this goal, the idealized MIEC material shown in Fig. 4 has not yet been realized. All of the current materials are lacking in one or more of the critical material requirements. An alternative approach to seeking one multifunctional material is to utilize an increasing number of materials, each one meeting one or more requirements. As with the Ni-YSZ anode for H_2 and CO fuels, the combination of materials meets all of the anode



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 4 Simplified schematic of an idealized MIEC-YSZ based anode operating with CH_4

requirements and can provide high performance. However, more solid phases lead to increased complexity of the anode mechanism and increased complexity in device fabrication to ensure sufficient contact between phases.

In reality, it is unlikely that the catalyst will be 100% selective and it is also difficult to perceive that the required eight-electron charge transfer reaction will occur at a single point on the catalyst surface. Pathway 1 in Fig. 4 represents an idealized view of the target process. There are numerous alternative reaction pathways, one of which is highlighted by the steps 2–4 in Fig. 4. In this case, total oxidation of methane occurs as in step 1 above; however, the product CO_2 can react further in a dry reforming reaction with more CH_4 to form CO and H_2, step 3. These products then undergo further electrocatalytic oxidation to CO_2 and H_2O, step 4. This is, again, a representative oversimplification that serves only to demonstrate the complexity of the system. The number of possible reaction pathways for fuel oxidation is very large as products, intermediates, and reactants interact and compete for active heterogeneous reaction sites, and undergo homogenous gas-phase reactions at SOFC temperatures. This complexity only increases as the C-number of the hydrocarbon fuel is increased. Fully unraveling all of these possible reactions and their contributions toward determining fuel cell performance is very challenging.

Some insight into the dominant reaction in the anode can be provided by the cell OCP. The theoretical OCP for an electrochemical cell is given by the Nernst equation. For a cell conducting oxygen anions, this is given by:

$$E = -\frac{RT}{2F}\ln\frac{fO_2^{\frac{1}{2}(a)}}{fO_2^{\frac{1}{2}(c)}} \tag{3}$$

The fugacity of oxygen at the cathode is usually well determined as the cathode is typically exposed to air. For the anode side of the cell, the fugacity will be set by the reaction equilibria between the fuel and oxidation products. This is well defined for H_2 fuel as only one reaction can occur:

$$H_2 + \frac{1}{2}O_2 \leftrightarrow H_2O \tag{4}$$

Substituting this equilibrium expression for the anode fugacity and assuming an ideal gas yields:

$$E = E^0 - \frac{RT}{2F}\ln\frac{P_{H_2O}^{(a)}}{P_{H_2}^{(a)}P_{O_2}^{\frac{1}{2}(c)}} \tag{5}$$

where E^0 is the standard potential for the fuel oxidation reaction at the temperature of interest. The theoretical OCP for humidified H_2 at 700°C is 1.122 V and an H_2-fuelled cell should generate this well-defined OCP. Oxygen (or air) leaks through poor cell sealing or residual porosity in the electrolyte should be avoided as they will increase the oxygen fugacity at the anode and decrease the cell OCP. Electronic conductivity within the electrolyte will also reduce the OCP due to the presence of an electrical short circuit within the electrolyte. However, a well-constructed and sealed YSZ-based SOFC will generate the theoretical OCP for hydrogen fuel as YSZ is an almost pure ionic conductor.

The equivalent theoretical anode oxygen fugacity for hydrocarbon fuels is not set by a simple reaction equilibrium due to the complex reaction chemistry of these fuels. There are a large number of elementary steps, potential reaction pathways, and reaction intermediates. The most desirable reaction is the total oxidation of CH_4 to CO_2 and H_2O, Eq. 2. This provides both the highest OCP value and represents full utilization of the fuel. If the catalyst is 100% selective for this

reaction, the anode oxygen chemical potential will be set by this reaction equilibrium:

$$CH_4 + 2O_2 \leftrightarrow CO_2 + 2H_2O \quad (6)$$

In this case, the cell OCP would be fully defined by substituting this equilibrium expression into Eq. 3 to yield:

$$E = E^0 - \frac{RT}{8F} \ln \frac{P_{H_z}^{2(a)} P_{CO_2}^{(a)}}{P_{CH_4}^{(a)} P_{O_2}^{\frac{1}{2}(c)}} \quad (7)$$

However, as discussed with reference to Fig. 4, the complexity is large, even for CH_4. CH_4 can undergo total oxidation to CO_2 and H_2O, Eq. 2, or partial oxidation to CO, H_2, and/or H_2O, Eq. 8. CH_4 may also undergo surface catalyzed steam or dry reforming with product CO_2 or H_2O, Eqs. 9 and 10; react to form graphitic carbon and H_2, Eq. 11; or undergo gas phase cracking and free radical polymerization to deposit tar-like residue. All of these reaction equilibria and chemical species combine to set the anode oxygen fugacity and OCP. Furthermore, while it is desirable to selectively catalyze the electrochemical total oxidation of CH_4, this reaction involves the transfer of eight electrons. It is difficult (if not impossible) to ensure that each of these steps will occur electrocatalytically without desorption and further non-electrochemical reaction of partial oxidation products:

$$CH_4 + 3O^{2-} \rightarrow CO + 2H_2O + 6e^- \quad (8)$$

$$CH_4 + H_2O \rightarrow CO + 2H_2 \quad (9)$$

$$CH_4 + CO_2 \rightarrow 2CO + 2H_2 \quad (10)$$

$$CH_4 \rightarrow C + 2H_2 \quad (11)$$

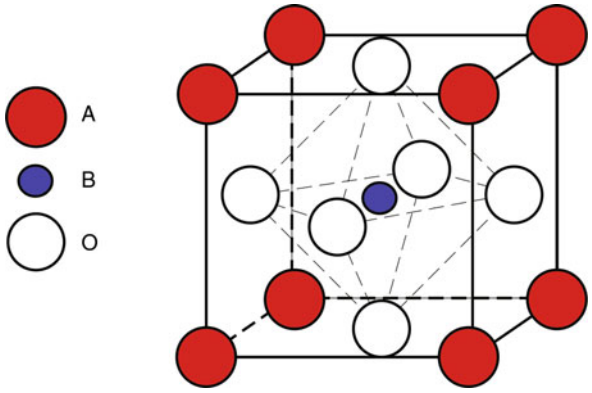
The difficulty in defining fugacity, expected variation of anode catalyst reaction selectivity for heterogeneous reactions, and influence of anode gas residence time on the prevalence of homogeneous reactions leads to significant variation in reported cell OCP for hydrocarbon fuels. This also leads to some debate over the definition of "direct" oxidation or utilization. In this entry direct utilization refers to the direct feeding of a hydrocarbon fuel to the anode in the absence of significant diluent, H_2O, CO_2, or O_2, as adopted by McIntosh and Gorte [1].

One barrier to enhancing electrocatalytic activity is the current lack of knowledge regarding the electrocatalytic fuel oxidation mechanism in the anode. There are very few studies in the literature that seek to understand the mechanistic details of this process. In large part this is due to the complexity of the anode system and the relatively extreme anode operating environment. A classical surface science approach is not applicable to the study of SOFC systems. Many of the materials of interest have variable oxygen stoichiometry that controls the ion-electron conductivity and, most likely, catalytic properties of the material. Since oxygen stoichiometry is set by the gas-phase pO_2, temperature, and local electrochemical environment, measurements made under vacuum or at low temperature outside of a working electrochemical system are difficult, if not impossible, to relate to the performance and mechanisms of working SOFC. While one of the most intriguing and challenging aspects of SOFC systems is the unique environment in which the materials are placed, this environment presents considerable challenges to the researcher. For example, while rates of non-electrocatalytic reactions on a powder catalyst sample can be measured, it is not currently possible to accurately probe surface intermediate species under an applied potential. While some groups have attempted to relate externally measured activity to SOFC performance [6, 18–22] and have made steps toward measurements under realistic operating conditions [20], the current lack of in-situ spectroscopic tools is a significant barrier toward rational design of new SOFC electrocatalysts .

Structures and Defect Chemistry

The majority of new materials for SOFCs are perovskite structured oxides of general form $ABO_{3-\delta}$ [23]. The ideal perovskite structure is a cubic close-packed ABO_3 structure where the B-site cation sits within the octahedral interstices, Fig. 5. This structure is very flexible toward cation composition and tolerates large substitution fractions on either cation site. The Goldschmidt factor, a ratio of A, B, and O ionic radii, is often utilized to predict if a metal oxide will crystallize into the perovskite structure [24]. The A site of the commonly utilized perovskites is typically occupied by La, Ca, Sr, or Ba. The B site is typically a transition metal. Other structures investigated include double perovskites, apatites, and fluorites.

There are a number of excellent reviews and thorough treatments of the defect chemistry of oxides



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 5 Schematic of the cubic perovskite structure

available in the literature, e.g., [25]. The brief overview below serves only to introduce the reader to primary concepts as a basis to the discussion in the rest of this entry. The following discussion is restricted to point defect models as these are predominant for most of the materials discussed.

Oxygen ion transport in these mixed oxides occurs primarily via oxygen anion hopping between oxygen vacancies. Cubic crystal structures are often considered desirable as all of the oxygen sites are geometrically equivalent. Structures with ordered oxygen vacancies may provide rapid transport through vacancy "channels" but perpendicular transport is frustrated by the ordered structure. In the absence of vacancy trapping or significant structural changes, ionic conductivity increases with increasing density of vacant oxygen sites and increasing mobility. Oxygen vacancies in these oxides are typically denoted as a deviation from the fully occupied lattice, the δ in 3-δ. Oxygen vacancies are formed either by reduction of a cation and/or by aliovalent doping.

Considering the very general system $A^{3+}B^{3+}O_{3-\delta}{}^{2-}$, a fully stoichiometric oxygen sublattice, $\delta = 0$, will exist while the combined charge on the A and B-sites is 6+, balancing the negative charge on the three O^{2-} anions. At elevated temperature and/or reduced gas-phase oxygen partial pressure, pO_2, the B-site transition metal cation may be reduced with the concomitant formation of oxygen vacancies. In Kröger–Vink notation:

$$2B_B^x + O_O^x \leftrightarrow 2B'_B + V'_O + \frac{1}{2}O_2 \qquad (12)$$

In addition to providing a mechanism for oxygen anion conduction, these oxygen vacancies can act as intrinsic electronic donors, with electronic conductivity occurring via a hopping mechanism between B and B′ sites.

An alternative approach to generate oxygen vacancies is aliovalent substitution of the element on the A or B site. For example, upon substituting X^{2+} for A^{3+}, the difference in charge can be compensated by two different mechanisms. The substitution can be compensated for through a change in the oxidation state of B to maintain a fully occupied oxygen sublattice:

$$XO + A_A^x + B_B^x + \frac{1}{4}O_2 \rightarrow X'_A + B^{\bullet}_B + \frac{1}{2}A_2O_3 \qquad (13)$$

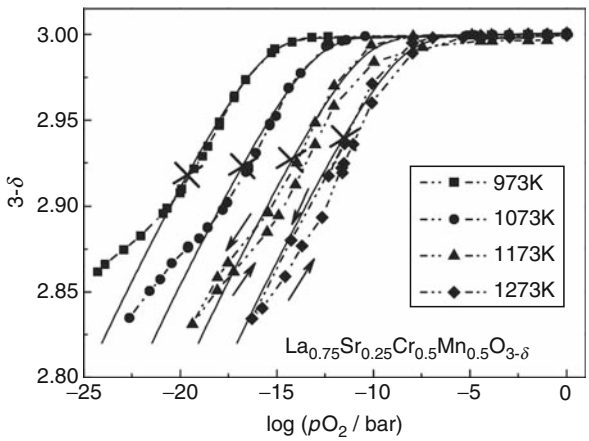
In this case the number of electronic carriers is increased. Alternatively, the B site can maintain a constant charge and the A-site substitution is compensated through the formation of oxygen vacancies:

$$\begin{aligned} &2XO + 2A_A^x + 2B_B^x + O_O^x \\ &\rightarrow 2X'_A + 2B_B^x + V_O^{\bullet\bullet} + A_2O_3 \end{aligned} \qquad (14)$$

The dominant compensation mechanism, either cation reduction or vacancy formation, is a function of pO_2 and temperature. Thus by selecting materials with differing reducibility in the anode environment, electronic, ionic or mixed ionic, and electronic conductors can be created. An example of oxygen nonstoichiometry data for a candidate anode material is shown in Fig. 6. Note that the oxygen stoichiometry decreases with decreasing pO_2 and increasing temperature within the typical anode pO_2 and temperature range.

Electronic Conductivity for Direct Hydrocarbon Anodes

In order to minimize ohmic losses in the anode, the composite structure should have an electronic conductivity greater than 100 S/cm [4]. One option is to replace Ni with a metallic phase that is inert toward catalyzing carbon formation. This is the approach adopted by Gorte, Vohs and coworkers who developed direct hydrocarbon anodes utilizing Cu as the electronic conductor [1]. While this approach is very promising, the list of usable metals is very short as most transition metals either catalyze carbon formation or are prohibitively expensive.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 6 Oxygen non-stoichiometry of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-\delta}$ as a function of pO_2 and temperature (Reprinted from [26] with permission from Elsevier)

An alternative approach, and one with a wide possible material palette, is to develop stable and highly conductive oxides. The primary challenge is to meet the required target conductivity for both the pure material and a composite. When mixed with an electrolyte, the overall composite electronic conductivity will be considerably lower. Most attempts at this seek to boost the density of p-type carriers in the base lattice by substitution with lower valence cations. Charge compensation upon doping can be either electronic or ionic, and *n*- or *p*-type conductors can be pursued. Details specific to each material are discussed within the relevant sections in this entry.

Poisoning and Stability

Homogenous gas-phase cracking and subsequent free radical polymerization of hydrocarbon fuels at SOFC operating temperatures leads to the formation of tar-like carbon in the anode compartment [27]. This can lead to blocking of anode active sites but can also lead to improvements in anode electronic conductivity [27, 28]. This is an inherent property of hydrocarbon fuels and, as such, is a hurdle toward long-term operation for any hydrocarbon fuelled SOFC. Methods to control this carbon deposition include utilizing more stable hydrocarbons (typically CH_4), reducing cell operating temperature, co-feeding diluent or steam, or develop anodes that aid in removing these deposits by oxidation, steam reforming, and/or dry reforming with product H_2O and/or CO_2. It is essential to distinguish between this tar-like carbon formation and the catalytic formation of graphitic carbon that occurs on Ni-based anodes. This catalytic graphite formation is a property of Ni metal and must be avoided through the design of new anode materials and composites.

In addition to these undesirable side reactions, impurities in the feed must also be considered. Sulfur can be a significant poison, with its exact concentration and form dependent on the source of hydrocarbon fuel. While sulfur can be removed by preprocessing, at the time of writing (2010), the US Environmental Protection Agency requires a refinery average sulfur content of gasoline fuel of 30 ppm or lower, with a per gallon cap of 80 ppm. Pacific Gas and Electric Company in California allows a total sulfur content of 17 ppm. A number of additional impurities must be considered if coal is to be used as the carbon feedstock, including P, K, Ni, and Cl. While their influence on Ni-based anodes has been considered [29–31], this issue has not been addressed for other anode materials.

Copper-Ceria

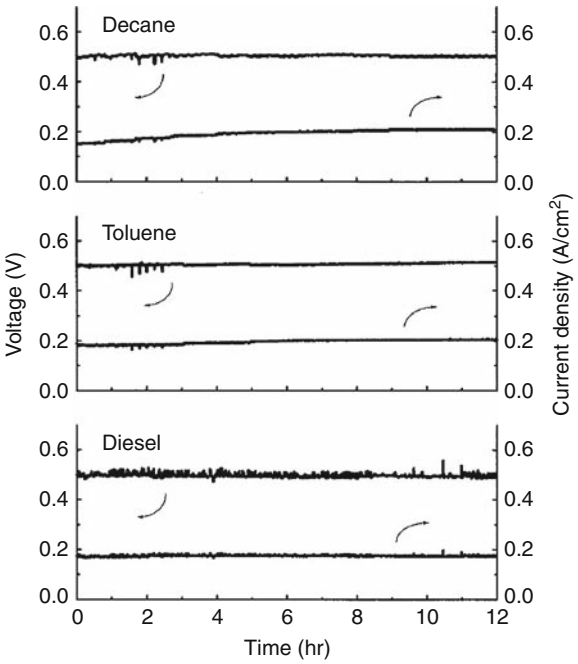
Gorte, Vohs, and coworkers at the University of Pennsylvania pioneered the use of Cu-CeO_2-YSZ composite anodes for direct hydrocarbon utilization [1]. Their approach was to separate the anode material requirements between three materials: Cu to provide electronic conductivity without catalyzing graphitic carbon formation, CeO_2 to catalyze the oxidation of hydrocarbon fuel, and YSZ to provide ionic conductivity. These anodes cannot be manufactured utilizing the co-firing approach used for Ni/YSZ due to the relatively low melting point of CuO (1,336°C). Instead the anodes are fabricated by tape-casting a slurry of YSZ with pore formers [32]. This layer is then co-fired with the electrolyte to form a dense YSZ electrolyte supported on a thicker porous YSZ "skeleton." Following addition of the cathode, the CeO_2 and Cu phases are added via wet infiltration of metal nitrate salts followed by low temperature calcination (450°C). This infiltration approach to anode manufacture enables the incorporation of anode materials at lower temperatures. This expands the potential materials palette to materials that

would either melt, sinter heavily, decompose, or react with YSZ during high temperature sintering of the YSZ phase. This is a flexible technique that has been demonstrated for both anode and cathode manufacture [33–35]. The largest potential drawback is the time-consuming procedure of adding salt solutions, firing, and repeating until sufficient weight-loading is reached. The associated increases in cell performance must be weighed against increased manufacturing costs.

The ability to utilize hydrocarbon fuels is clearly an advantage to this approach. Cu-CeO_2-YSZ anodes have been demonstrated to operate with hydrogen fuel and via direct utilization with a range of hydrocarbon fuels from CH_4 to synthetic diesel fuel. Figure 7 shows operation of these cells with a variety of hydrocarbon fuels.

Conductivity

The YSZ in the Cu-CeO_2-YSZ composite anode is present as the primary ionic conductor. Since YSZ is a pure electrolyte, the electronic conductivity of the anode must be provided by a second phase. Although reduced CeO_2 has an electronic conductivity of greater than 1 S/cm at 1,000°C [37], the electronic conductivity in a 15 vol% CeO_2-YSZ composite was only 0.019 S/cm at 700°C in H_2 [38]. While this low electronic conductivity of CeO_2/YSZ composites is sufficient in a thin anode functional layer [39], a thicker anode structure requires higher electronic conductivity in order to minimize ohmic losses. As such, an additional phase is required. Unlike Ni, Cu does not catalyze the formation of graphitic carbon deposits in the anode, thus enabling the cell to operate with dry hydrocarbon fuel feed.

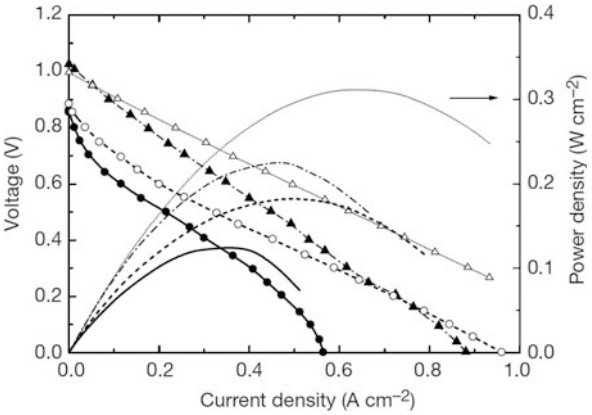


Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 7 Cell voltage and current density versus time at 700°C for a cell with 10/20/70 wt% CeO_2/Cu/YSZ anode with various hydrocarbon fuels (Reproduced with permission from [36]. Copyright 2001, The Electrochemical Society)

It should also be noted that the electronic conductivity of a porous YSZ-based composite electrode will be significantly lower than that of the pure electronic conducting phases due to the porous and electron blocking YSZ. The loading of the electronic conductor required in the anode composite is dictated by percolation considerations. The electronic conductivity of a 15 vol% Cu-YSZ composite can be as high as 1,190 S/cm [38]. It should be noted that 15 vol% is much lower than the >40 vol% Ni typically utilized in Ni-YSZ anodes. This decrease in required metal loading is due to the difference in fabrication procedure. The Ni-YSZ powder processing leads to random mixing of the two phases and >40 vol% Ni is required to ensure percolation [3, 40]. In contrast, wet infiltration of Cu salt solution into a preformed porous YSZ structure leads to coating of the pore walls and formation of contiguous conducting pathways at lower volumetric loading. The electronic conductivity of the composite is thus a strong function of Cu loading, and the infiltration must be controlled to form the desired Cu structure [38].

Although Cu does not catalyze graphite formation, most hydrocarbon fuels undergo gas-phase free radical polymerization when fed undiluted into the anode compartment at SOFC operating temperatures. The long-chain and aromatic [22] products of this polymerization lead to tar-like deposits within the Cu-CeO_2-YSZ anode structure. While excessive deposition of these compounds may lead to blocking of the active sites within the anode, their electronic conductivity has a beneficial influence by reducing ohmic losses within the anode. McIntosh et al. [27] examined

the performance of SOFCs with varying Cu loading before and after exposure to dry n-C_4H_{10} at 700°C for 30 min. It was found that the initial ohmic resistance of cells with 10 wt% CeO_2 and either 5, 10 or 20 wt% Cu in H_2 fuel at 700°C was significantly higher than that calculated for the YSZ electrolyte (i.e., there was a significant ohmic resistance in the electrodes). For the cell with 20 wt% Cu, the ohmic resistance decreased dramatically upon exposure to n-C_4H_{10}, reaching the value predicted for the YSZ electrolyte. A cell with 10 wt% CeO_2 and 30 wt% Cu showed no change in ohmic resistance, matching that of the electrolyte before and after exposure to n-C_4H_{10}. This shift in ohmic resistance was ascribed to bridging gaps between isolated Cu particles with conducting tar-like carbon deposits. Indeed, further work [21] demonstrated operation of Cu-free cells utilizing only these deposits as the predominant electronic conductor.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 8 SOFC performance curves for cells with Cu-ceria composite anode. The cell had a 60-μm electrolyte, and data are shown for the following fuels: *filled circles*, *n*-butane at 700°C; *open circles*, *n*-butane at 800°C; *filled triangles*, H_2 at 700°C; and *open triangles*, H_2 at 800°C (Reprinted by permission from Macmillan Publishers, Ltd: Nature [43], copyright 2000)

Catalysis

While Cu can be utilized as a catalyst, it does not contribute significantly to the overall catalytic activity of the anode. This was verified by the low performance of Cu-only anodes, particularly in hydrocarbon fuels [22], and the identical performance achieved when Cu is replaced with catalytically inert bulk Au [41]. The role of CeO_2 as electrocatalyst for fuel oxidation was confirmed by replacing CeO_2 with other lanthanide oxides and comparing SOFC performance with the activity of the lanthanide toward fuel oxidation [22]. The cell performance tracked well with the n-C_4H_{10} light-off temperature of the lanthanide.

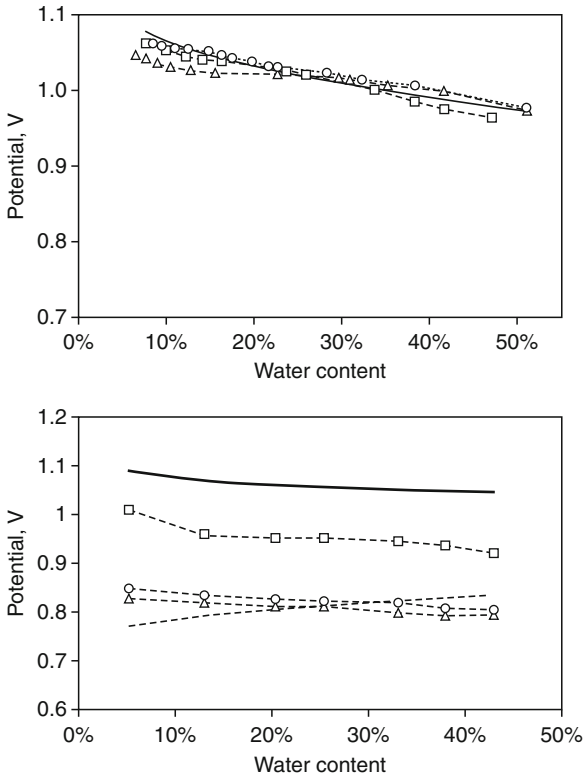
The performance of Cu-CeO_2-YSZ based cells (in terms of both cell OCP and maximum power density) is consistently lower for hydrocarbon fuels than for H_2. For example, Park et al. [42] reported cell OCP values of 1.05 and 0.9 V respectively for H_2 and CH_4 fuels at 700°C, with corresponding maximum power densities of 0.32 and 0.09 W/cm^2. A similar decrease is observed when comparing H_2 and n-butane fuel, Fig. 8. Kim et al. reported [36] a cell OCP of 1.1 V with current density at 0.5 V of 0.5 A/cm^2 at 700°C in H_2 fuel; these values decreased to 0.8 V and 0.2 A/cm^2 in 35 vol% toluene/N_2 fuel at the same temperature on the same SOFC.

Since the cathode and electrolyte are constant between tests, the electrocatalytic activity of the anode is the primary cause of this decrease. This limitation was partially overcome by utilizing precious metal dopants in the anode [21]. The addition of 1 wt% Pd added to a C-CeO_2-YSZ anode resulted in the same OCP, polarization resistance, and hence power density for CH_4 and H_2 fuels. The OCP in CH_4 increased from 1.0 V to 1.25 V and the maximum power density increased from 20 to 280 mW/cm^2 upon the addition of 1 wt% Pd to nominally identical anodes. This was attributed to the enhanced electrocatalytic activity of Pd-CeO_2 toward CH_4 oxidation compared to pure CeO_2. It should be noted that this cell utilized carbonaceous deposits as the electronic conducting component in the anode due to alloying between Pd and Cu. This result also suggests that care must be taken when utilizing precious metal current collectors to assess the performance of novel anodes. The current collector may inadvertently contribute significant catalytic activity to the anode.

While the influence of the electrocatalyst on cell power output is clear and trends in oxidation activity measured by traditional means track with cell

performance, very little is known about the fundamental reaction mechanism occurring in the anode. As discussed above, the thermal stability of hydrocarbons toward homogeneous cracking and free radical polymerization decreases with increasing carbon number. At SOFC operating temperatures, most hydrocarbons will readily undergo these gas-phase reactions in the hot areas of the fuel feed system and in the anode pores. Examining the OCPs reported for n-C_4H_{10} and longer-chain hydrocarbons on Cu-CeO_2 anodes finds that the OCPs are all very similar at around 0.8 V [28]. This suggests that the fuel species reaching the anode and the reactions occurring are similar for all these fuels, which makes sense when considering thermal cracking.

McIntosh et al. [21] studied the trend of OCP with CO_2 and H_2O content for H_2, CH_4, and n-C_4H_{10} fuels with CeO_2-C-YSZ anodes with and without the addition of Pd. For H_2, the trend of OCP with H_2O content for both anodes tracked the theoretical trend for total oxidation, as this is the only reaction equilibrium that can be established for H_2 fuel. For CH_4, Fig. 9, the trend for Pd-free anodes more closely follows the trend expected if CH_4 first undergoes reforming with the feed CO_2 and H_2O and it is the product H_2 that reacts at the anode. The trend line was calculated assuming the equilibrium amount of H_2 interacts with the anode. Upon addition of Pd, the measured OCP increased, with the trend line lying between the trend lines calculated for the reforming based mechanism of Pd-free electrodes and for direct CH_4 oxidation. This suggests that Pd-CeO_2 is active toward direct oxidation of CH_4. While an increase in OCP was measured upon addition of Pd for n-C_4H_{10} fuel, the increase was smaller. In addition, it was shown [27, 28] that the tar-like hydrocarbons formed by free radical polymerization are deposited in the anode structure of SOFCs with these longer-chain hydrocarbon fuels. It is likely that these deposits are the primary interaction species at OCP, even with precious metal catalyst dopants, but they may be removed from the active sites in the anode via oxidation with product H_2O under cell operation [28, 44].

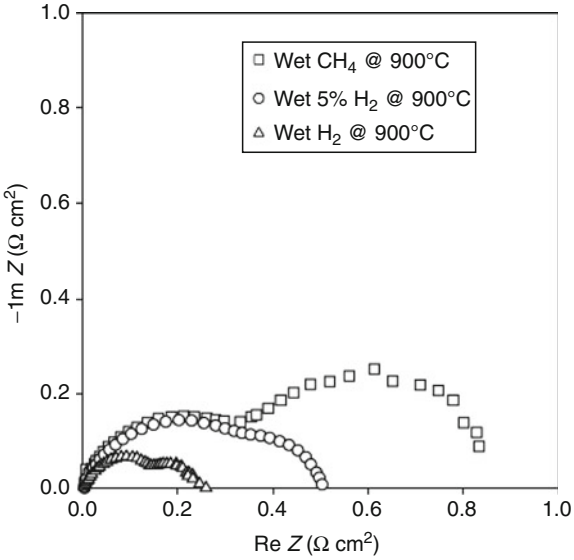


Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 9 OCP as a function of H_2O content in H_2-H_2O (*top*) feed and CH_4-CO_2-H_2O (*bottom*) feed with CH_4:CO_2 ratio of 8. The calculated Nernst potentials for total oxidation of fuel are shown as bold lines (both figures) while the dashed line assumes reforming to form H_2 (*bottom*). Experimental results are shown for: C-ceria-YSZ (*triangles*); Cu-ceria-YSZ (*circles*); and C-Pd-ceria-YSZ (*squares*) anodes (Reproduced with permission from [21]. Copyright 2003, The Electrochemical Society)

Poisoning and Stability

The sulfur tolerance of these anodes is primarily dictated by the thermodynamics of ceria sulfide and oxysulfide formation, as the sulfur concentration required for Cu_2S formation is significantly higher [45]. Kim et al. demonstrated that stable operation could be achieved if the sulfur level in the fuel is reduced to 100 ppmv S as thiophene in 5 mol% n-$C_{10}H_{22}$ – this is below the concentration predicted from thermodynamics for Ce_2O_2S formation. This study was followed by work of He et al. who demonstrated stable operation up to 450 ppmv H_2S in H_2 [46], significantly higher than the tolerance levels reported for Ni-based anodes [47].

As these cells operate at temperatures relatively close to the melting point of Cu (1,085°C), concerns have been raised regarding the thermal stability of Cu. While Cu will sinter at SOFC operating temperatures, the thermal stability of the Cu phase is shown to be a strong function of the infiltration technique and precursor. These parameters influence the distribution of Cu throughout the anode structure [38]. Alloying may be utilized to increase the melting point, although alloy metals and alloy level must be chosen carefully in order to avoid nullifying the benefits of Cu with regard to graphitic carbon formation. Kim et al. [48] demonstrated the use of Cu-rich Ni alloys in preventing graphitic carbon formation and showed 500 h of operation in dry CH_4 with 80–20 mol% Cu-Ni replacing Cu in the anode. Lee et al. [49] expanded this to include Cu-Co based anodes and proposed that segregation and subsequent surface coverage of the Co by a Cu layer provided stable operation with dry hydrocarbon fuels. With regard to the CeO_2 phase, doping with Zr has been shown to significantly enhance thermal stability [50].



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 10 Electrode impedance of an optimized $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_3$ anode at 900°C, in different humidified fuel gas compositions. The electrolyte contribution has been subtracted from the overall impedance (Reprinted by permission from Macmillan Publishers, Ltd: Nature Materials [52], copyright 2003)

Lanthanum Chromates

Lanthanum chromate ($LaCrO_{3-\delta}$) is a reasonable electronic conductor and is stable up to high temperatures (> 1,000°C) in both oxidizing and reducing ($pO_2 < 10^{-21}$ atm) environments [51]. Perhaps the most promising material in the $LaCrO_{3-\delta}$ family so far is $La_{0.75}Sr_{0.25}Cr_{0.50}Mn_{0.50}O_{3-\delta}$ (LSCM). Single phase LSCM anodes showed low anode polarization resistances, 0.26 and 0.87 $\Omega \cdot cm^2$ in humidified H_2 and CH_4, respectively, at 900°C [52], Fig. 10. High power densities, 0.86 and 0.48 W/cm^2, were observed for SOFC with LSCM/Cu composite anodes in 850°C H_2 and CH_4, respectively [53]. Note that electrolyte thickness, electrode geometry, and cathode material can dictate power density as much as anode material, and that power densities should only be compared between nominally identical cells.

As (La,Sr)CrO_3 itself is not catalytically active toward hydrocarbon oxidation [54], Cr has been partially substituted with a variety of elements, limited mostly to first-row transition elements (Ti, V, Mn, Fe, Co, and Ni) and Ru, in an effort to enhance catalytic activity [55–58]. These substitutions significantly impact the stability of the perovskites. (La,Sr)CrO_3 with 5–25 mol% of Ru [59] on the B site or with large amounts (20–50 mol%) of Co [56], Fe [60] or Ni [56, 61] substituting for Cr are unstable, as the dopant cations exsolute from the perovskite lattice under reducing conditions. $LaCrO_3$ substituted with Ti is stable in reducing and oxidizing conditions, but has poor catalytic activity for fuel oxidation [56], showing large polarization resistances when used as anodes [62]. This is likely due to the low redox activity of Ti. Finally, the introduction of significant amounts ($\geq$ 10 mol%) of V into the (La,Sr)CrO_3 lattice requires firing in highly reducing conditions [19]. It is likely that, as with LSV, V-substituted lanthanum chromates will not be stable in air [63].

The best substituent for Cr seems to be Mn, as LSCM shows stable performance in humidified CH_4 at 900°C [64], and has some catalytic activity for the total oxidation of hydrocarbons [65]. Although Zha et al. [66] reported that LSCM was stable in pure H_2 at 950°C ($pO_2 < 10^{-20}$ atm), Kawada found significant

quantities of the decomposition products $LaSrMnO_4$ and MnO in LSCM pellets after thermogravimetric measurements at lower temperatures and under less reducing conditions (900°C, $pO_2 = 10^{-20}$ atm) [26]. The same phases were encountered by other researchers after treating LSCM powder in humidified 20% H_2/N_2 at 800°C for 8 h [19]. These differences are expected to be due to variations in treatment times and synthesis procedures. The results indicate that LSCM may not be completely stable in pure fuel at very high temperatures. However, the actual pO_2 values in a working anode are expected to be higher than the pO_2 in the surrounding atmosphere due to the transport of oxygen through the SOFC, which could stabilize the anode material.

Conductivity

The total conductivity of $La_{0.75}Sr_{0.25}Cr_{0.50}Mn_{0.50}O_{3-\delta}$ was reported to be ~40 S/cm in air, and 1.5 S/cm in 5% H_2 at 900°C [64, 67] Fig. 11. While constant at high pO_2, total conductivity decreased sharply for $pO_2 < 10^{-10}$ atm [67], indicative of *p*-type conductivity [64, 68]. Positive values for the Seebeck coefficient under oxidizing conditions confirmed that conductivity in LSCM is *p*-type dominant [67, 69]. At low temperatures and in high pO_2 atm, the net negative charge created by the substitution of La^{3+} with Sr^{2+} is compensated by an increase in the average oxidation state of Mn, generating electron holes localized on Mn that are responsible for electronic conductivity [68], Eq. 15.

$$4SrO + 4La_{La}^{\times} + 4Mn_{Mn}^{\times} + O_2(g) \leftrightarrow 4Sr'_{La} + 2La_2O_3 + 4Mn_{Mn}^{\bullet} \quad (15)$$

The activation energy for the conductivity is on the order of 0.1 eV, indicative of conduction through a thermally activated polaron-hopping mechanism.

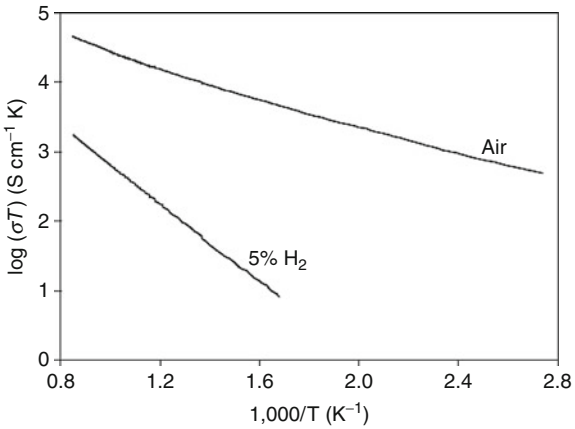
At elevated temperatures and in low pO_2 atm, Mn is reduced in average oxidation state and the net negative charge is now compensated by the formation of oxygen vacancies, Eq. 16, Fig. 6.

$$4Mn_{Mn}^{\bullet} + 2O_O^x \leftrightarrow 4Mn_{Mn}^x + 2V_O^{\bullet\bullet} + O_2(g) \quad (16)$$

The decrease in electron holes, evident from Eq. 16, results in a decrease in total conductivity in atmospheres with $pO_2 < 10^{-10}$ [67, 70]. Due to the generation of oxygen vacancies, this decrease is accompanied by an increase in oxygen ion conductivity, from $\sim 4\times10^{-5}$ S/cm at high pO_2 to $\sim 4 \times 10^{-4}$ at $pO_2 = 10^{-15}$, measured at 950°C. Although the ionic conductivity will be even higher in fuel atmospheres ($pO_2 \leq 10^{-21}$ atm), the conductivity under these conditions was found to remain dominated by electron holes [67].

The total conductivity of 1.5 S/cm in 5% H_2 is further reduced in the SOFC anode and by the use of composite porous GDC/LSCM or YSZ/LSCM anodes, to values of ~0.1 S/cm in humidified H_2 at 800°C for a YSZ/LSCM composite [71]. Although this value is much lower than the 100 S/cm suggested by Steele [4], thin YSZ/LSCM anodes with 1 wt% Ni show reasonable performance in H_2 nonetheless [34].

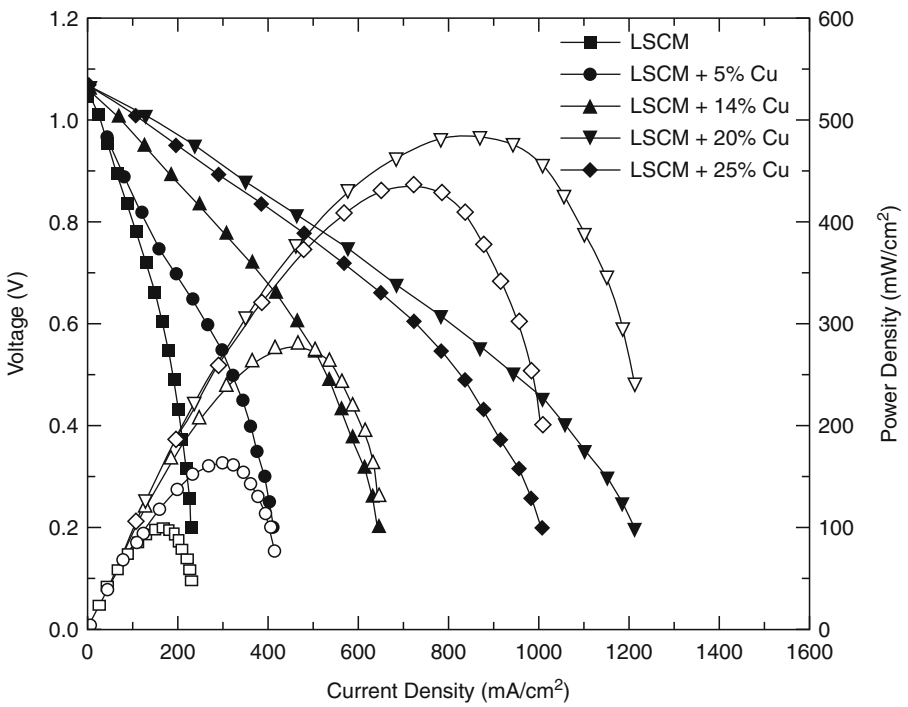
The substitution of La^{3+} with up to 50 mol% of Sr^{2+} has been attempted to enhance the electronic conductivity of LSCM [70]. However, the conductivity at low pO_2 is almost independent of the amount of Sr in the range of 0.20–0.30 mol% Sr [68], and the introduction of more than 0.25 mol% Sr did not significantly enhance electric conductivity. Furthermore, the presence of ≥ 30 mol% Sr resulted in higher solid state reactivity and decreased the compatibility with the YSZ electrolyte [70]. Increasing the amount of Mn on the B site was reported to enhance the conductivity both in air and reducing atmospheres, but decreased the stability at low pO_2 [66].



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 11 Temperature (T) dependence of total conductivity (σ) of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-\delta}$ in air and 5% H_2 (Reprinted by permission from Macmillan Publishers, Ltd: Nature Materials [52], copyright 2003)

Another method is the substitution of La^{3+} with alkaline earth elements other than Sr. Since the different alkaline earth elements are present in the 2+ oxidation state at the relevant conditions, the number of electron holes is not expected to change significantly for substitution with different alkaline earths. However, changes in lattice parameters and in the distance between the Mn centers, caused by differently sized A-site cations, could affect the electrical conductivity through changes in charge carrier mobility. The electrical conductivity was shown to change when substituting La with Mg and Ba, but this was mostly due to the formation of secondary phases [72]. Apart from Sr, only substitution with Ca resulted in a phase-pure perovskite, but without an increase in conductivity compared to LSCM [72]. Furthermore, the introduction of Ca is expected to result in the instability of LSCM under reducing conditions [73].

In practice, electronic conductivity has been increased by developing composite anodes of LSCM and an electronic conductor. Cu in particular has been used for the latter [6, 53], since it is an excellent electronic conductor and does not promote hydrocarbon cracking. Therefore, the large amounts of material required to achieve a percolation path can be added without carbon deposition during SOFC operation in hydrocarbon fuels. Although the electronic conductivity was not measured directly, Zhu measured a decrease in anode overpotential and a concomitant increase in SOFC power density with the addition of increasing amounts of Cu to LSCM [53], Fig. 12. The decrease in overpotential was mainly due to a decrease in the ohmic resistance of the SOFC. Since Cu replaced the catalyst LSCM, an optimum in Cu amount was expected. Indeed, this optimum was observed at 20 wt % Cu, at which point the current density was increased more than fourfold, compared to the pure LSCM anode. With the addition of more Cu, the low-frequency impedance part, ascribed to fuel-related oxidation processes, increased significantly. This may be due to blocking of the active area on the LSCM within the anode.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 12
Voltage and power density for single cells with different LSCM + Cu composite anodes at 850°C in dry CH_4 (Reprinted from [53] with permission from Elsevier)

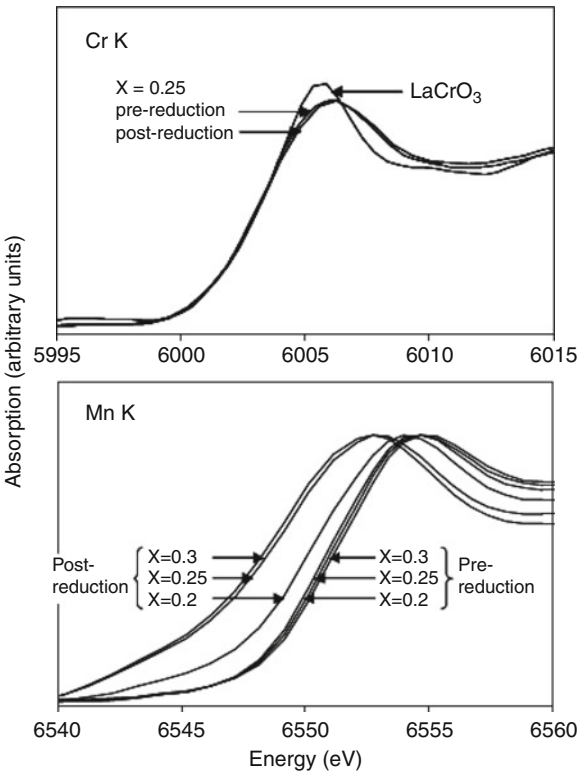
To our knowledge, a systematic enhancement of the ionic conductivity has not been attempted. Instead, SOFC anodes typically use a composite of LSCM with an electrolyte material, such as YSZ [71] or ceria [74]. Care has to be taken when attributing increases in cell performance to enhanced ionic conductivity, as ceria has some total oxidation activity and could function as a catalyst. Also, at least part of the increase in performance is likely to be structural, since the porous YSZ or GDC scaffold creates a large surface area for the oxidation reaction to take place.

Catalysis

Although the exact reaction mechanism is unclear, general concepts regarding the fuel oxidation reaction on LSCM have been described in the literature. Yamazoe and Teraoka pointed out that high-temperature oxidation reactions on perovskites typically occur through a reduction-oxidation cycle of the catalysts, otherwise known as a Mars-van Krevelen-type (MvK) mechanism [75], with the B-site elements serving as the redox centers. Studies suggest that this is the case for LSCM.

The active center of LSCM was identified by reducing air-annealed LSCM in fuel atmosphere, and comparing the energy levels of Cr and Mn before and after the reduction using X-ray Adsorption Near Edge Structure (XANES) analysis, Fig. 13. While the Cr K edge energies remained constant, the energy levels were different for Mn before and after reduction, attributed to a difference in oxygen coordination. This indicates that the average oxidation state of Mn decreases during reduction [68], identifying Mn as the active redox center. This is further supported by the poor catalytic activity of $LaCrO_{3-\delta}$ toward hydrocarbons, compared with the higher activity of $LaMnO_{3-\delta}$ [54].

Initial catalytic tests indicated that LSCM catalyzes the oxidation of hydrocarbon fuels, although it also has some activity toward dry reforming [65]. Compared to the conversion of CH_4 in a blank reactor, LSCM enhanced the reaction rate of CH_4, and increased the selectivity toward total oxidation. Even under relatively oxygen-lean conditions (CH_4:O_2 gas mixtures of 4:1), LSCM promoted the total oxidation of CH_4 to CO_2, up to 800°C. At 850°C, the CO_2/CO selectivity was still 98.6% in favor of CO_2. No significant steam reforming activity was measured.



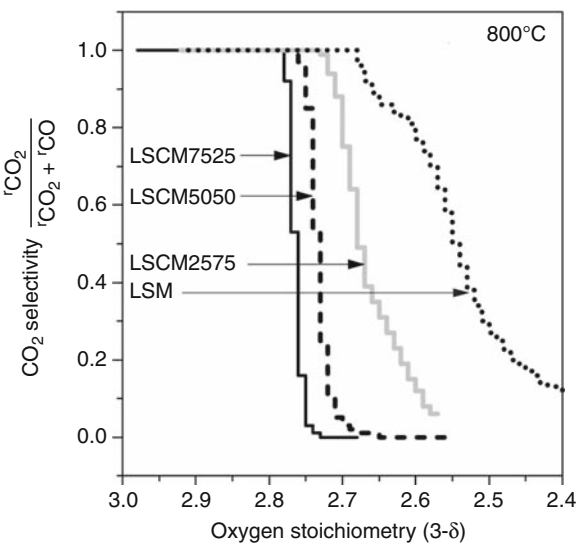
Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 13 Cr and Mn K XANES spectra for $La_{1-x}Sr_xCr_{0.5}Mn_{0.5}O_{3-\delta}$ (Reprinted from [68] with permission from Elsevier)

The experiments described above were performed by co-feeding oxygen and CH_4, resulting in a $pO_2 \sim 10^{-1}$ atm, much higher than the pO_2 of $\sim 10^{-20}$ present at a working SOFC anode. Furthermore, the measured reaction rates are not differential rates as they were determined for 100% conversion. In order to determine the oxidation rate and the selectivity for total and partial oxidation of CH_4 under conditions similar to those of operating anodes, van den Bossche and McIntosh developed a pulse-type reactor [20]. In this set-up, pulses of CH_4 (or H_2) are fed to a reactor containing oxidized powder catalyst. No oxygen is co-fed with the fuel; instead, the fuel is oxidized utilizing oxygen from the oxide catalyst surface and bulk. Inert gas is fed between the short pulses of hydrocarbon fuel to allow re-equilibration of the surface oxidation states. This is necessary to prevent the measurement of a bulk oxygen diffusion limited rate. The pO_2 above the catalyst during a hydrocarbon pulse is that of pure fuel,

and the oxygen source is the same as in the anode. Thus the CO_x production rates and selectivity are those of the fuel cell. Furthermore, since each pulse titrates a small amount of oxygen from the lattice, multiple pulses and knowledge of the original stoichiometry enable measurement of rates and selectivity as a function of oxygen stoichiometry.

Pulse experiments on $La_{0.75}Sr_{0.25}Cr_{1-x}Mn_xO_{3-\delta}$ ($x = 0.25 - 1$) indicated that CH_4 oxidation rates increased with increasing Mn substitution and decreased with decreasing oxygen content. Both observations are in agreement with a modified MvK mechanism with Mn as the redox center. The oxygen stoichiometry (3-δ) of the perovskite influenced product selectivity. At high lattice oxygen content, $\delta \approx 0$, total oxidation was the preferred reaction for all Mn/Cr ratios, with 100% of CH_4 reacting to CO_2, Fig. 14. When less lattice oxygen was present, CO production dominated. The formation of CO was suggested to occur through the partial oxidation of CH_4, rather than secondary dry reforming.

Attempts have been made to link this shift in reaction rate and selectivity with oxygen stoichiometry to observed shifts in SOFC anode polarization resistance. At open circuit conditions at 700°C with CH_4 fuel, the LSCM oxygen stoichiometry is low as the material is at equilibrium with the fuel gas. This appears to promote cracking as the main reaction, with correspondingly high anode polarization resistance, ~12.3 Ω·cm^2 [6]. Upon application of current, it is suggested that the oxygen ion flux locally increases the oxygen stoichiometry of the perovskite. The pulse reactor studies suggest that this should shift the reaction mechanism toward total oxidation of the fuel and increase the activity. This is observed as a decrease in anode polarization (to 1 Ω·cm^2) and a shift toward CO_2 production with increasing SOFC current density.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 14 Selectivity toward total oxidation of methane at 800°C. The black solid line is LSCM with 75 mol% Cr and 25 mol% Mn (LSCM7525), dashed line is LSCM5050, gray solid line is LSCM2575 and dotted line is LSM (Reprinted from [20] with permission from Elsevier)

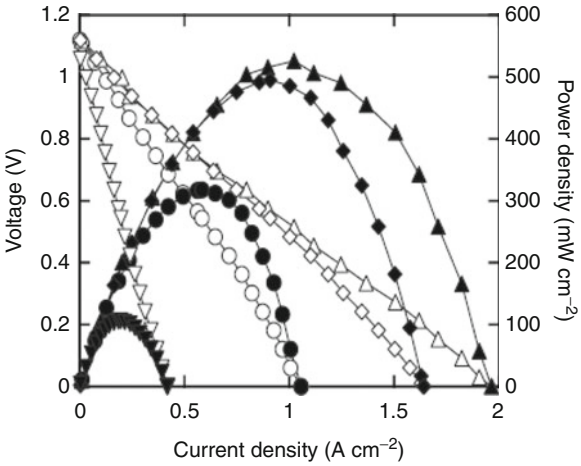
While these results are interesting, SOFC with LSCM anodes show significantly lower power outputs in C_4H_{10} and CH_4, when compared with H_2 fuel [6]. Since the change in fuel only influences anode conditions, the anode activity toward oxidation of hydrocarbons appears to be a performance-limiting process for these anodes. This was further illustrated by electrochemical measurements on SOFCs with dense LSCM films [18]. The OCP values measured for these SOFCs with CH_4 on the anode were comparable to the values measured in inert gas (He), ~0.06 V. It was suggested that the kinetics for CH_4 oxidation are slow, and that the contact time of the fuel on the relatively smooth surface was too short to set the CH_4 reaction equilibrium. The addition of small amounts of Pd as a catalyst to the anode resulted in reasonable OCP values, ~0.87 V, proving that the catalytic performance of LSCM toward CH_4 is low and must be improved.

To enhance catalytic activity, the B site of LSCM has been substituted with a third element that is suggested to be more active for hydrocarbon oxidation than Mn. Ni and Fe are attractive choices as their respective lanthanum perovskites show CH_4 conversion at lower temperatures than $LaMnO_3$ [54]. Also, Ni is used as the oxidation catalyst in traditional Ni/YSZ anodes. Although large amounts of Co, Fe, or Ni substitution are unstable [56, 61], small amounts of Ni (10 mol%) were suggested to be kinetically stable in the $LaCrO_3$ lattice [73]. Furthermore, the use of relatively small amounts of Co, Fe, or Ni, even if exsoluted, is not expected to lead to the mechanical failure that is caused by large amounts of especially Ni, as carbon formation and volume changes upon redox cycling will be limited.

Van den Bossche and McIntosh replaced 10 mol% of Mn with Co, Fe, or Ni and found that CH_4 oxidation rates were greatly enhanced on LSCMNi and LSCMCo, up to an order of magnitude compared to LSCM. This occurred only after the exsolution of the Co and Ni metals. The degree of carbon formation for LSCMNi and LSCMCo was similar to LSCM [19], suggesting that these compositions can be used as redox-stable, direct oxidation anodes.

Instead of introducing small amounts of catalytic elements into the perovskite, catalysts have also been successfully added to the anode as a separate phase. Again, Ni is the primary focus, as are precious metals such as Pd, Pt, and Rh. By adding a relatively small amount (4 wt%) of Ni to a GDC/LSCM (80% Cr) anode, Liu et al. [74] were able to increase the performance of SOFCs from ~50 to 80 mW/cm^2 in 750°C CH_4 (the relatively low power output resulted partly from a thick electrolyte). As the small amount of Ni was not expected to significantly increase the electronic conductivity, the increase in power output is mainly due to the enhancement in catalytic activity of the GDC/LSCM-Ni anode. Although the composite anodes with 4% Ni generated a small amount of carbon deposits in C_3H_8 fuel under open circuit conditions, no noticeable carbon formation occurred in working SOFCs. This was in contrast to composites with large amounts of Ni, which generated significant amounts of carbon, even under operating conditions. Although the SOFC power output of the GDC/LSCM anode with 4% Ni dropped slightly after a first redox-cycle, performance was stable during the next three cycles. Long-term electrochemical stability studies are required to determine the stability of this composite anode.

Using a similar method, Kin et al. obtained a power output of ~500 mW/cm^2 in humidified H_2 at 700°C utilizing YSZ/LSCM anodes with 0.5–1 wt% of either Pd, Rh or Ni added as a separate phase, Fig. 15. These values are to be compared to the power output of ~100 mW/cm^2 for YSZ/LSCM without metals added. The addition of 1 wt% Fe to YSZ/LSCM also increased power output to 400 mW/cm^2 [34]. An enhancement in SOFC performance with the addition of Pd was found by other researchers as well, but not to the extent of a fivefold increase [53]. The smaller impact may be due to other limiting processes, such as low ionic conductivity.



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 15 V-I polarization curves (open symbols) and power densities (filled symbols) in humidified H_2 at 700°C for cells having anodes with 45 wt% LSCM in YSZ, using various catalysts: (▼) no catalyst, (●) with 5 wt% ceria, (♦) with 0.5 wt% Pd, and (▲) with 5 wt% ceria and 0.5 wt% Pd (Reproduced with permission from [34]. Copyright 2009, The Electrochemical Society)

Pd and Rh may be prohibitively expensive for large-scale SOFC applications, but Ni or even Fe would be suitable. Long-term fuel cell tests using (sulfur-rich) CH_4 are required to assess the sulfur tolerance and sintering resistance of Ni in the YSZ/LSCM-Ni anode.

Stability

The LSCM anode is chemically compatible with the commonly used electrolytes, such as YSZ, up to 1,300°C [64] and LSGM, up to 1,100°C [76]. Values for the thermal expansion coefficient (TEC) of LSCM in air varied from 8.9 to 10.8 μK^{-1} at low temperatures ($<$ ~500°C), and from 10.1 to 12.7 μK^{-1} at high temperatures, up to 950°C [64, 67]. These TEC values match well with those for YSZ (10.8 μK^{-1}), LSGM (11.1 μK^{-1}) and GDC (13.5 μK^{-1}) [8]. The change in TEC in LSCM with temperature is likely due to a change in space group from rhombohedral R-3 C to cubic Pm-3 m, occurring gradually over the range of 500–1,100°C in air [77]. As the associated change in volume is minimal, ~1% [64], this is not expected to cause instability during cell heating or

cooling. Furthermore, since the Pm-3 m phase is also observed for reduced LSCM, stresses due to redox cycling are expected to be small. Indeed, the difference in volume between the reduced and oxidized unit cells is only 1.7% at room temperature [77], while the difference in linear expansion between LSCM in air and in reducing atmospheres, $pO_2 = 2\times10^{-19}$, is limited to ~0.2% at 700°C [67].

The operation of SOFCs with LSCM anodes in sulfur-rich H_2 leads to a decrease in cell performance in a short amount of time. Liu reported a decrease in current density of 50% when exposing LSCM to 10% H_2S/H_2 at 950°C [66]. The decrease occurred over a period of 2 h, after which the performance stabilized. Ten percent H_2S is much higher than the typical ppm levels of sulfur present in natural gas and represents attempts to utilize H_2S as a fuel. However, even in H_2 with 50 ppmv H_2S, performance was observed to decrease significantly from 0.46 W/cm^2 to 0.09 W/cm^2, again in 2 h, at 850°C and constant current density of 625 mA/cm^2 [53]. The large deterioration of LSCM is due to La and Mn reacting with sulfur to form La_2O_2S, MnS, and α-MnOS [66]. Since Mn is the reactive site for fuel oxidation, the formation of these compositions is likely to have an adverse effect on anode performance.

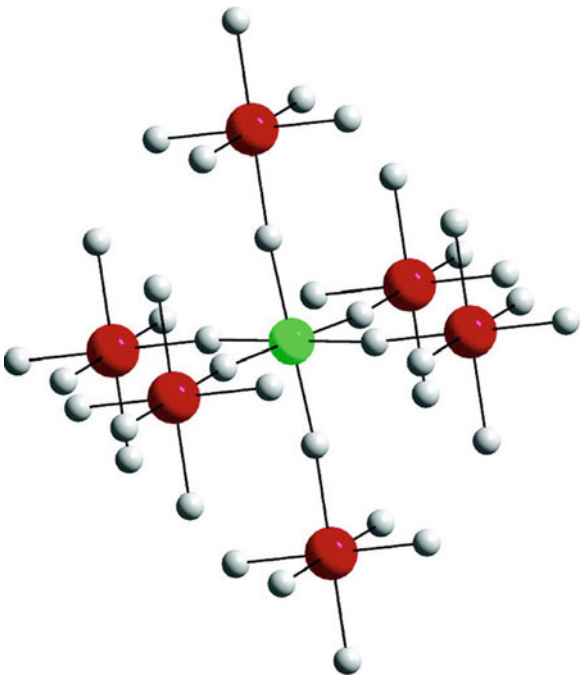
There is no easy solution to decrease the susceptibility of LSCM to sulfur poisoning. Since LSV anodes have been shown to have excellent sulfur resistance [78], operating better in sulfur-rich fuel than in fuel without sulfur, it has been attempted to introduce 10 mol% V on the B site of LSCM. However, no phase-pure material could be obtained [19].

Interestingly, the use of sulfur-rich CH_4 instead of sulfur-containing H_2 led to an initial increase in performance on LSCM, compared to CH_4 without sulfur [56]. After 4–6 h though, the performance in 0.5% H_2S/CH_4 had stabilized to values similar to those in pure CH_4. This observation is consistent with the much smaller and slower decrease of power density of SOFC with LSCM anodes in 0.5% H_2S/CH_4, as compared to H_2S/H_2 mixtures [79].

Double Perovskites

Double perovskites are described by the general formula $A_2B'B''O_{6-\delta}$, and can be interpreted as a doubling of the regular perovskite $ABO_{3-\delta}$. The elements on the B′ site are different from those on the B″ site, and the two sites alternate to form a sublattice with a rock-salt structure. In a perfectly ordered lattice, every B′ ion would be coordinated by six perovskite unit cells with B″ ions and vice versa. This is illustrated in Fig. 16 for $Sr_2MgMoO_{6-\delta}$ (SMMO); the Mo ion (green) is surrounded by 6 Mg ions (red). The ordering is caused by large differences in the radii and charges between the B′ and B″ site ions. For example, in SMMO, the B′ site is occupied by the small Mg^{2+} ion, while the B″ site hosts the large and highly charged Mo^{6+} ion [80]. Although a regular perovskite can have two different elements on the B site as well, it is this particular B-site structure that classifies these materials as double perovskites.

The most investigated double perovskite anode material is SMMO. It has reasonable electrical conductivity, ~10 S/cm in pure H_2 at 800°C, good activity for direct CH_4 oxidation, and operates well in



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 16 Mo (*green*) surrounded by six next-nearest-neighbor Mg (*red*) in a perfectly B-site ordered material; oxygen ions are white and strontium cations are omitted for clarity (Reprinted in part with permission from [80]. Copyright 2007 American Chemical Society)

sulfur-containing fuels [81]. The power output of an SOFC with a 300-μm-thick $La_{0.8}Sr_{0.2}Ga_{0.8}Mg_{0.2}O_{3-d}$ electrolyte, $SrCo_{0.8}Fe_{0.2}O_{3-d}$ cathode and SMMO anode reached 0.84 W/cm^2 in dry H_2 and 0.44 and 0.34 W/cm^2 in dry and wet CH_4, respectively, at 800°C [81], Fig. 17. Note that these values were obtained using Pt current collectors, which should be avoided as they contribute toward the catalytic activity of the anode [82], significantly increasing the performance for hydrocarbon fuels.

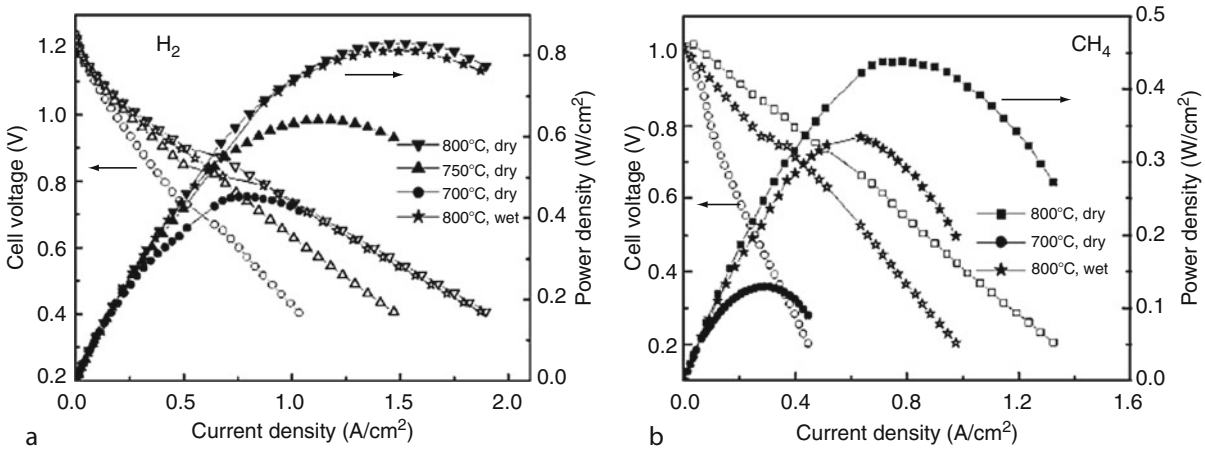
Conductivity

The electrical conductivity of SMMO was found to be ~1–4 S/cm in H_2/Ar and 10 S/cm in H_2, at 800°C [83, 84]. The slight differences in the values possibly result from differences in synthesis methods. The conductivity decreases with increasing pO_2, indicative of *n*-type conductivity. The charge carriers are electrons generated by the reduction of Mo(VI) to Mo(V) to compensate for the charge imbalance created by the formation of oxygen vacancies in the material [84]. No studies have yet been conducted to optimize the electrical conductivity.

Catalysis

The catalytic activity of SMMO is thought to result from the Mo ion. Whereas the Mg ion keeps a 2+ charge, Mo is reduced to some extent from predominantly Mo(VI) in air atmospheres, to a mixture of Mo (VI) and Mo(V) in reducing atmospheres [85]. The presence of a redox center would allow for the oxidation of fuel molecules according to a MvK mechanism, as suggested for the oxidation of CH_4 on LSCM [20]. Only a small amount of Mo(VI) is reduced to Mo(V), resulting in an oxygen non-stoichiometry of ~0.04–0.06 mol O/mol SMMO for samples reduced at 1,000–1,200°C in 5% H_2 [80, 84, 86]. The limited reduction of Mo(VI) could be due to the instability of Mg in a conformation with less than six surrounding oxygen ions. This is confirmed by the decreased B-site order after reduction, indicating that Mg ions relocate to maintain a sixfold oxygen coordination [80].

To increase the electrocatalytic activity of Mo-based double perovskites, Mg has been replaced by a number of transition metals. SOFC anodes have been prepared with Sr_2CoMoO_6, Sr_2FeMoO_6, Sr_2MnMoO_6 and Sr_2NiMoO_6, but none of those compounds were stable. SMMO with Co, Ni and Zn on the B′ site are easily synthesized in air, but are unstable in 5% H_2 at 800°C and above, due to exsolution of the metals [87, 88]. Higher initial power outputs were indeed observed for SOFCs made with Sr_2CoMoO_6 anodes, compared to similar SOFCs with SMMO anodes, but performance quickly deteriorated, likely due to the decomposition of



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 17

SMMO/LDC/LSGM/SCF cell voltage and power density as a function of current density in dry and wet: (**a**) H_2; and (**b**) CH_4. The open symbols represent the cell voltages while the closed symbols represent the power densities (Reproduced with permission from [83]. Copyright 2006, The Electrochemical Society)

SCMO. Sr_2FeMoO_6 and Sr_2MnMoO_6 are stable in reducing atmospheres, but unstable in >400°C air [88, 89].

More recently, the maximum oxygen non-stoichiometry in SMMO was successfully increased through partial substitution of Mo(VI) by Nb(V), without compromising the stability of the double perovskite [85]. The increase in oxygen non-stoichiometry suggests that more Mo^{6+}/Mo^{5+} redox centers are created, which are expected to increase both oxygen ion conductivity and fuel oxidation rates.

Stability

As suggested above, a significant challenge in the development of double perovskites is the synthesis of phase-pure, redox-stable compositions. The synthesis of SMMO, for example, is successful in air, but requires high temperatures, >1,200°C, to remove $SrMoO_4$ impurities [90]. The firing temperature can be brought down to 900°C if a 5% H_2 atm is used, but the use of such a reducing atmosphere could complicate large-scale SMMO production.

Because of the existence of $SrMoO_4$ impurities in air-fired SMMO, it is likely that the same impurities are formed upon treatment of H_2-synthesized SMMO in high temperature air. Depending on the thermal and chemical expansion properties of this phase, these impurities could result in a decrease in performance with redox cycling of SMMO. At room temperature in air, SMMO readily forms carbonates, such as $SrCO_3$, resulting in surface degradation and $SrMoO_4$ formation [84].

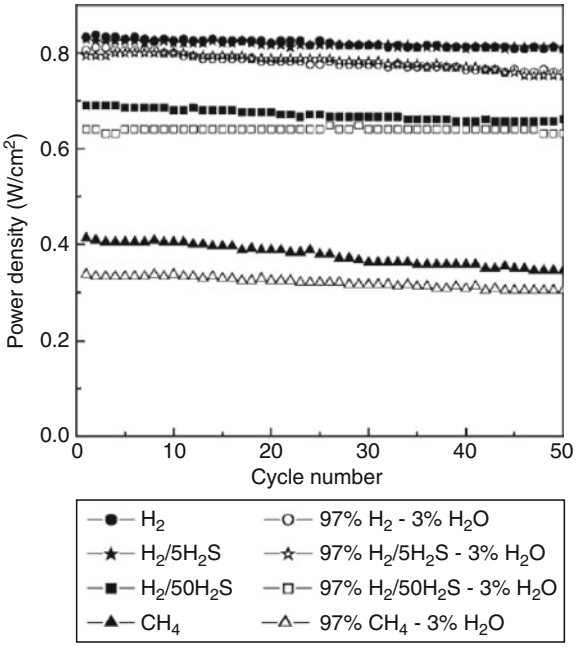
At elevated temperatures in reducing atmospheres, SMMO partly decomposes to MgO, SrO, and Mo. The temperature at which decomposition starts is subject to debate. In 5% H_2, decomposition temperatures have been reported as low as 900°C [80] up to 1,000°C [90] and even 1,200°C [83]. It has been suggested that the decomposition temperature is influenced by the preparation method, with the high temperatures utilized during air-firing or solid-state synthesis possibly leading to the evaporation of the volatile Mo, decreasing SMMO stability [90]. SMMO synthesized according to an EDTA-complexation route and fired at low temperatures in 5% H_2 is typically found to be stable up to 1,000°C in 5% H_2, at which point small amounts of the decomposition products appear. SOFC applications however run most efficiently on pure or humidified fuel, atmospheres that are more reducing than 5% H_2. The use of pure fuel could therefore lead to decomposition of the SMMO anode at temperatures lower than 900–1,000°C, which presents a problem for SOFC operating at 700–850°C.

Another potential issue is the chemical interaction of SMMO with the commonly used electrolytes YSZ and LSGM. When treated in 1,000°C air for 24 h, SMMO reacts with YSZ to form large amounts of $SrMoO_4$ and $SrZrO_3$, and it reacts with LSGM to form $(La,Sr)Ga_3O_7$ [84]. These reaction products have inferior properties compared to the reactants and are therefore to be avoided. In the case of LSGM, this has been done by using a thin ceria-based layer between anode and electrolyte [83]. The thermal expansion of SMMO, ~11.7–12.7 μK^{-1} [83] is compatible with values for YSZ (10.8), LSGM (11.1) and GDC (13.5).

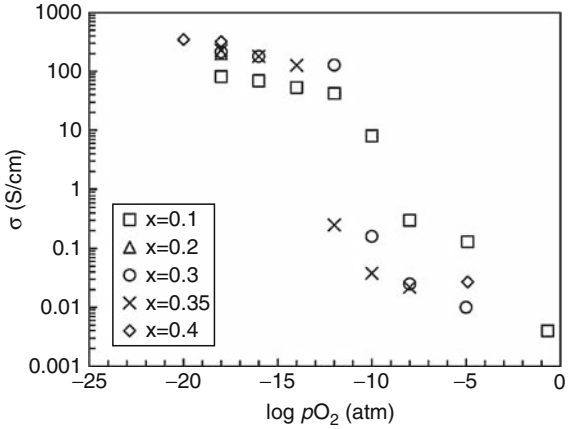
SOFC with SMMO anodes have shown stable performance in dry H_2 fuel with 5 ppmv H_2S. The initial power output of fuel cells in 5 ppmv H_2S/H_2 fuel is similar to that in pure H_2 fuel, Fig. 18, and the cells show a similar decrease in power output, ~3% after 200 h of operation in both fuels at 800°C [83]. When, under otherwise similar conditions, the concentration of H_2S is raised to 50 ppmv, the initial power output of the cell was reduced to ~80–85% of its value in H_2 and P_{max} decreased ~5% in 200 h.

Strontium Titanates

Pure $SrTiO_3$ is a simple cubic perovskite that has a good thermal expansion match with YSZ and is stable under reducing conditions. However, undoped $SrTiO_3$ has insufficient electronic conductivity for an anode material [91]. In contrast, donor-doped (*n*-type) $SrTiO_3$ has attracted considerable interest due to its high electronic conductivity under reducing conditions, while maintaining a good thermal expansion match, and resistance to both coking and sulfur poisoning [92, 93]. Unfortunately, the performance of pure STO-based anodes is low, requiring additional catalytic materials to realize acceptable power densities [94]. As with the other oxides discussed in this entry, the conductivity and stability of this material is



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 18 Maximum power density P_{max} at 800°C versus cycle number in dry and wet H_2, H_2/H_2S, and CH_4 (Reproduced with permission from [83]. Copyright 2006, The Electrochemical Society)



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 19 Electrical conductivity as a function of pO_2 for $La_xSr_{1-x}TiO_3$ sintered at 1650°C in H_2. Conductivity is measured at 1,000°C from low to high pO_2 (Reprinted from [96] with permission from Elsevier)

dominated by its defect chemistry. There are a number of articles that fully discuss the defect chemistry, for example [91]. This discussion is restricted to the essential points relevant to the use of STO as an SOFC anode.

Conductivity

There are two routes to doping of $SrTiO_3$: doping of trivalent cations onto the Sr site, and/or doping of pentavalent cations onto the Ti site. The most common trivalent dopants are La^{3+} and Y^{3+}. La^{3+} has high solubility, with only limited distortion of the lattice parameter at high dopant levels (>40%) [95]. Charge compensation upon doping can occur via electronic or ionic compensation, depending on the temperature and pO_2. At low pO_2, the compensation mechanism is electronic via reduction of Ti^{4+} to Ti^{3+} to form $Sr_{1-x}La_xTi'_xTi_{1-x}O_3$:

$$\frac{1}{2}La_2O_3 + Sr_{Sr}^{x} + Ti_{Ti}^{x} \rightarrow La_{Sr}^{\cdot} + Ti'_{TI} + SrO + \frac{1}{4}O_2 \quad (17)$$

A similar mechanism occurs for Y-doped STO under reducing conditions [97]. Assuming that dopant carriers dominate, the number of n-carriers is then directly correlated to the dopant level [98]; however, the correlation is not 100% as some electronic defects are associated with the donor atom. The creation of high concentrations of electronic defects leads to high conductivity. For example, Marina et al. reported conductivities for $La_{0.3}Sr_{0.7}TiO_3$ of over 100 S/cm at low pO_2, Fig. 19 [96].

At high pO_2, the compensation mechanism switches from electronic to cation vacancy compensation (or self compensation), where the donor charge is compensated by the formation of Sr vacancies [98, 99]. This is accompanied by the formation of a secondary Sr-rich phase. The nature of this phase is not fully determined and is suggested to be either an $Sr_{n+1}Ti_nO_{3n+1}$ phase within the matrix [100] or a separate SrO phase [101]. A separate phase is denoted below simply for clarity:

$$Sr_{Sr}^{x} + 2Ti'_{Ti} + \frac{1}{2}O_2 \leftrightarrow SrO + V''_{Sr} + 2Ti_{Ti}^{x} \quad (18)$$

This alternate charge compensation mechanism removes electronic carriers and thus greatly decreases the electronic conductivity of the material [98]. Indeed, the very high conductivities reported for La-doped

STO are only created upon sintering at high temperatures under very reducing conditions [92, 93, 96, 102].

Upon exposure to higher pO_2 at SOFC temperatures, the conductivity decreases slowly [96]. This is observed in Fig. 19 as the significant decrease in conductivity for all of the samples as the pO_2 is increased above $\sim 10^{-14}$ atm. The slow kinetics associated with this may be related to slow cation diffusion [101] and/or low oxygen mobility in these materials [103]. Slow reoxidation kinetics are beneficial to application in the SOFC anode where accidental and occasional exposure of the anode material to air may be expected. Of course, this assumes that a suitable and cost-effective cell synthesis procedure can be derived to initially form these highly conductive states. Furthermore, slow reoxidation may indicate slow surface oxygen exchange and low hydrocarbon oxidation activity.

Nb is the most common pentavalent dopant [104–108]. As with the trivalent A-site dopants, the additional charge is compensated by electronic carriers at low pO_2 leading to high electronic conductivity. These carriers are generated by reduction of Ti^{4+} to Ti^{3+}:

$$\frac{1}{2}Nb_2O_5 + 2Ti_{Ti}^{x} \rightarrow Nb_{Ti}^{\bullet} + Ti'_{Ti} + TiO_2 + \frac{1}{4}O_2 \quad (19)$$

Nb remains in the pentavalent oxidation state although there is evidence that Nb substitution improves the overlap between Ti atomic orbitals. As with the trivalent doped materials, conductivity decreases upon oxidation due to a shift toward ionic compensation via the formation of Sr vacancies. However, it has been reported that Nb-doped materials can provide adequate conductivity and are very slow to reoxidize [107].

Irvine et al. suggested deliberate introduction of Sr vacancies through the fabrication of A-site deficient materials with the A-site deficiency level set to compensate for the charge introduced through donor doping [103, 109]. While these materials show good electronic conductivity, recent work has suggested that A-site deficiency leads to precipitation of secondary Ti-rich phases [93, 110]. It should be noted that these secondary phases are not always observable by XRD but are observed upon SEM/EDX analysis.

Transition metal acceptor doping of A-site deficient $Sr_{0.85}Y_{0.1}Ti_{0.95}M_{0.05}O_{3-\delta}$ (M = V, Mn, Fe, Co, Ni, Cu, Zn, Mo, Mg, Zr, Al, Ga) leads to a decrease in total conductivity compared to the undoped composition. This is due to the acceptor dopant off-setting the influence of the donor dopant (Y^{3+} in this case). The ionic conductivity of these materials increases upon acceptor doping but it is still ~6 orders of magnitude lower than the electronic conductivity [111].

The ionic conductivity of doped STO is quite low although A-site deficient $(Y_{0.08}Sr_{0.92})_{1-x}TiO_{3-\delta}$ has been shown to have an ionic conductivity comparable to YSZ at 700°C; however, this occurs at the expense of lowering the electronic conductivity [112]. It may not be desirable to introduce significant oxygen ion conductivity as the low mobility of oxygen anions may play a significant role in preventing reoxidation of the material and concomitant loss of electronic carriers [96, 103]. The majority of SOFC anode studies utilize composites of doped STO and YSZ, where STO provide electronic conductivity and YSZ provides ionic conductivity [92].

Catalysis

Unfortunately, while able to provide sufficient electronic conductivity, the electrocatalytic activity of doped-$SrTiO_3$ is low [105]. As such, secondary catalysts are required to provide sufficient SOFC performance as cells without additional catalyst provide very low power density. For example, Lee et al. demonstrated a dramatic increasing in performance with H_2 fuel upon addition of 5 wt% CeO_2 and 0.5 wt% Pd. The power density increased from less than 20–780 mW/cm^2 at 800°C with no change in OCP [35]. Pd doping without CeO_2 has also been shown to enhance activity [94].

Avoiding precious metals, Fu et al. demonstrated a similar enhancement, realizing a two order of magnitude decrease in polarization resistance with an yttria-substituted $SrTiO_3$-YSZ composite upon infiltration of 5 vol% Ni [113]. Yang et al. also utilized Ni surface doping but did so in conjunction with Co doping of $Sr_{0.88}Y_{0.08}TiO_3$, finding an enhancement of activity attributed to the reducibility of Co within the lattice [114].

Only limited studies have discussed the activity of these materials toward hydrocarbon utilization. Vincent et al. showed that the performance of cells with $La_{0.4}Sr_{0.6-x}Ba_xTiO_3$ ($0 \leq \times \leq 0.2$) based anodes

in CH_4 fuel was more than one order of magnitude lower than in H_2 fuel between 800 and 950°C [115]. The performance did increase slightly with an increase in Ba doping level but, perhaps most intriguingly, a more substantial improvement was observed upon the addition of H_2S to the feed gas. A role for Ba in enhancing oxygen surface exchange rates was confirmed, although with the caveat that it was measured under oxidizing conditions, when the surface is doped with Sr, Ba, or Ca [116]. Addition of CeO_2 to an $Sr_{0.88}Y_{0.08}O_3$-based anode significantly reduced the anode polarization resistance with both H_2 and CH_4 fuel [117, 118], likely due to enhanced catalytic activity of the composite.

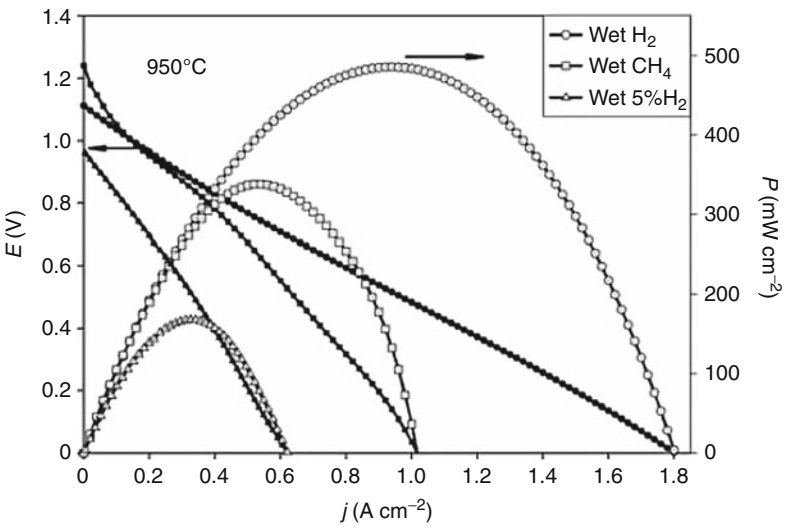
Doping with a more reducible Mn cation to form $La_{0.4}Sr_{0.6}Ti_{0.4}Mn_{0.6}O_3$ leads to a decrease in electronic conductivity although it may slightly increase catalytic activity toward hydrocarbons. The resulting cell shows a low OCP in wet CH_4 of only 0.86 V at 856°C with a total cell R_P of ~1.4 $\Omega \cdot cm^2$ compared with an OCP of 1.05 V and total R_P of ~0.36 $\Omega \cdot cm^2$ in wet Ar/H_2 [119]. The difference in polarization resistance is not as dramatic as shown for Mn-free compositions.

Ruiz-Morales et al. took a unique approach by utilizing multiple dopants to control the oxygen stoichiometry and thus disrupt defect ordering to generate the single-phase perovskite anode material $La_4Sr_8Ti_{11}Mn_{0.5}Ga_{0.5}O_{37.5}$ [120]. The resulting cell generated 500 mW/cm^2 with H_2 fuel and 350 mW/cm^2 with CH_4 at 950°C, Fig. 19. The maximum power density and OCP in CH_4 fuel were both significantly higher than with 5% H_2 fuel, Fig. 20, suggesting that homogenous cracking of CH_4 to form H_2 was not the dominant reaction mechanism and that this material was active toward CH_4 oxidation.

$SrTiO_3$ based anodes have been shown to be quite tolerant to H_2S poisoning [121], operating with up to 1% H_2S [122] and up to 50 ppm H_2S when surface doped with Pd [94].

Other Materials of Interest

Pyrochlore structured oxides have received some attention for use in SOFC anodes, in particular the series $Gd_2Mo_xTi_{2-x}O_7$. SOFCs made with 250-micron thick YSZ electrolytes, LSCM cathodes, and $Gd_2Mo_{0.6}Ti_{1.4}O_7$ anodes have shown a stable power output of 340 mW/cm^2 in 950°C 10% H_2S/H_2, and an anode polarization resistance of ~0.23 $\Omega \cdot cm^2$ at OCP [123]. The substitution of Ti in the electrolyte $Gd_2Ti_2O_7$ with large amounts of Mo increases the total conductivity to ~70 and 25 S/cm in 1,000°C 5% H_2 for 70 and

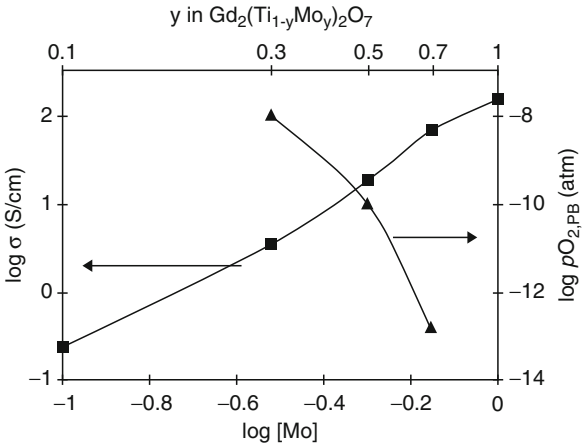


Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 20
Fuel cell performance plots at 950°C of a cell with four-layer optimized $La_4Sr_8Ti_{11}Mn_{0.5}Ga_{0.5}O_{37.5}$ anode, 330 μm YSZ electrolyte, and $La_{0.8}Sr_{0.2}MnO_3$ cathode. Fuels humidified with 2.3% H_2O, cathode is in dry O_2 (Reprinted by permission from Macmillan Publishers, Ltd: Nature [120], copyright 2006)

50 mol% Mo respectively. The ionic conductivity under these conditions remains high, ~0.1 S/cm for 50 mol% Mo [124].

However, the introduction of large amounts of Mo is detrimental to the stability of $Gd_2Ti_2O_7$. Compositions with 70 mol% Mo are unstable and exist only in a very narrow pO_2 range, 10^{-13}–10^{-15} atm at 1,000°C. This range is extended for lower Mo contents, but even at 30 mol% Mo, compositions are not stable at 1,000°C in $pO_2 > 10^{-8}$ atm [124], Fig. 21. Sprague reported increased redox stability when co-substituting Ti with Mn instead of just Mo, but this has only been shown for small amounts of Ti substitution, $\leq$ 20 mol% [125].

More recently, the pyrochlore $Yb_{0.96}Ca_{1.04}TiNbO_7$ has been suggested as an anode material [126]. Based on $Yb_2Ti_2O_7$, the substitution of Yb with Ca increases the ionic conductivity [127], while the introduction of Nb greatly enhances the total conductivity, to 9 S/cm at 800°C in 5% H_2/N_2. The conductivity is n-type, and is therefore expected to increase further in more reducing atmospheres. In addition, $Yb_{0.96}Ca_{1.04}TiNbO_7$ is stable up to 1,450°C in air and 1,350°C in 5% H_2/N_2. The main drawback of the material appears to be slow oxygen ion exchange, which can be slightly enhanced with further substitution on the B site, for example, with Mn, Cr or Mg/Mo [126].



Direct Hydrocarbon Solid Oxide Fuel Cells. Figure 21 Electrical conductivity at $pO_2 = 10^{-18}$ bar and the pyrochlore phase boundary at high pO_2, $pO_{2,PB}$, vs. Mo concentration in $Gd_2Ti_2O_7$ at 1,000°C (Reprinted from [124] with permission from Elsevier)

Fluorite anodes, not including ceria, are typically based on YSZ and have been developed to ensure good thermo-mechanical and chemical compatibility with the widely used YSZ electrolyte. A purely ionic conductor, YSZ depends on the substitution of Zr with more reducible elements, such as Ti, to introduce electronic conductivity. The reduction of Ti^{4+} to Ti^{3+} in $Y_{0.18}Zr_{0.73}Ti_{0.09}O_{2-\sigma}$ (YZT) generates electronic charge carriers, but significant amounts of Ti^{3+} are only present at high temperatures and low pO_2. Even at 1,000°C and $pO_2 < 10^{-20}$ atm, the total conductivity is low, ~0.06 S/cm [128]. An additional drawback of substitution with Ti is the trapping of oxygen vacancies on Ti ions, reducing the ionic conductivity to ~0.04 S/cm.

By raising the amount of Ti to 18 at-%, the maximum amount that dissolves in the YSZ lattice, the electrical conductivity of $Y_{0.15}Zr_{0.67}Ti_{0.18}O_{2-\delta}$ is increased to 0.2 S/cm (at 930°C and $pO_2 = 10^{-20}$ atm), while the ionic conductivity is further reduced [129]. Introduction of Sc_2O_3 into YZT (ScYZT) slightly increased the electronic conductivity to 0.14 S/cm at 900°C [130]. SOFCs with ScYZT anodes were tested in 900°C 5% humidified H_2 (using Pt as electrode current collectors) and showed an estimated power output of 35 mW/cm^2 and an anode polarization resistance of ~5.5 Ω·cm^2 at OCP, using a 2 mm thick YSZ electrolyte and a Pt cathode [131].

Among the *Tungsten Bronzes*, formula $A_{0.6}BO_3$, the Nb-based $Sr_{0.2}Ba_{0.4}Ti_{0.2}Nb_{0.8}O_3$ is the best candidate for SOFC anodes. This composition is redox stable and has reasonable electrical conductivity, 3.4 S/cm at 930°C in 5% H_2/Ar [132]. Other compositions are unstable or have lower conductivity [132, 133]. $Sr_{0.2}Ba_{0.4}Ti_{0.2}Nb_{0.8}O_3$ has a much lower TEC than the electrolytes that are commonly used, ~6.7 μK^{-1}, and reacts with YSZ at 1,200°C. Also, the power output for symmetrical SOFCs made with this compound are low, with impedance data showing Warburg behavior, suggesting poor oxygen ion exchange and low oxygen ion conductivity [134].

Vanadates in the series $La_{1-x}Sr_xVO_3$ demonstrate sufficient electronic conductivity [63], and reasonable electrocatalytic activity at 1,000°C [135], coupled with stability toward sulfur poisoning [78]. Aguilar et al. demonstrated operation with 5% H_2S in both H_2 and N_2 [135]. In the latter case, the LSV-based anode utilized H_2S as the fuel source. Unfortunately, LSV is

only stable at low oxygen partial pressure, for example, $pO_2 < 10^{-17}$ atm at 800°C [63]. It transforms to the apatite phase $Sr_3V_2O_8$ [63] or $Sr_2V_2O_7$ [78] upon exposure to higher pO_2. This transformation is irreversible at SOFC operating temperatures [63].

Future Directions

As discussed in this entry, a number of novel materials and composites have been proposed as potential anodes for direct hydrocarbon solid oxide fuel cells. While many are promising, a commercially viable solution has not yet been found. The discussion in this entry is deliberately framed around the concepts of ionic and electronic conductivity, electrocatalysis, and stability. It is essential for future researchers to address all of these topics when discussing new materials. The schematic in Fig. 4 represents both the complexity of the problem and the simplicity that could potentially be achieved if a material meeting all of these requirements can be found.

One particular challenge for the field as a whole is the measurement of these properties under realistic conditions. Experimental tools to probe the surface or bulk of these materials in situ at the length scales relevant to the anode processes are not currently available. While significant consideration has been given to the bulk transport properties of potential materials, very little is known of the surface chemistry under realistic SOFC conditions. It is also necessary to develop techniques that can measure and quantify the electrocatalytic activity of these materials under operating conditions. While electrochemical impedance techniques can, in principle, provide a wealth of information, the fitting of such data can be ambiguous and it is not a replacement for a direct measurement of a fundamental property. This is a major hurdle that must be overcome if highly active and selective catalytic materials are to be developed for this application.

In combination with this increased knowledge of surface chemistry, a detailed knowledge of the material bulk is essential. It is clear from the discussion in this entry that the target materials set will be a multicomponent oxide with potentially complex cation and anion lattice structures. When material properties depend upon cation configuration, e.g., the double perovskites, or oxygen stoichiometry, these properties must be accurately determined in a working cell.

The barriers to direct hydrocarbon SOFCs will only be overcome through a combination of new material development and more detailed experimental systems that directly probe fundamental parameters on working SOFC electrodes.

Bibliography

Primary Literature

1. McIntosh S, Gorte RJ (2004) Direct hydrocarbon solid oxide fuel cells. Chem Rev 104:4845–4865
2. Minh NQ (2004) Solid oxide fuel cell technology–features and applications. Solid State Ionics 174:271
3. Minh NQ (1993) Ceramic fuel-cells. J Am Ceram Soc 76: 563–588
4. Steele BCH, Middleton PH, Rudkin RA (1990) Material science aspects of SOFC technology with special reference to anode development. Solid State Ionics 40–41:388
5. Tao SW, Irvine JTS (2004) Discovery and characterization of novel oxide anodes for solid oxide fuel cells. Chem Rec 4:83–95
6. Bruce MK, van den Bossche M, McIntosh S (2008) The influence of current density on the electrocatalytic activity of oxide-based direct hydrocarbon SOFC anodes. J Electrochem Soc 155:B1202–B1209
7. Adler SB (2004) Factors governing oxygen reduction in solid oxide fuel cell cathodes. Chem Rev 104:4791
8. Jacobson AJ (2009) Materials for solid oxide fuel cells. Chem Mater 22:660
9. Tsipis EV, Kharton VV (2008) Electrode materials and reaction mechanisms in solid oxide fuel cells: a brief review. J Solid State Electrochem 12:1367–1391
10. Brett DJL, Atkinson A, Brandon NP, Skinner SJ (2008) Intermediate temperature solid oxide fuel cells. Chem Soc Rev 37:1568–1578
11. He HP, Hill JM (2007) Carbon deposition on Ni/YSZ composites exposed to humidified methane. Appl Catal, A 317:284–292
12. Gupta GK, Hecht ES, Zhu H, Dean AM, Kee RJ (2006) Gas-phase reactions of methane and natural-gas with air and steam in non-catalytic regions of a solid-oxide fuel cell. J Power Sources 156:434–447
13. Walters KM, Dean AM, Zhu H, Kee RJ (2003) Homogeneous kinetics and equilibrium predictions of coking propensity in the anode channels of direct oxidation solid-oxide fuel cells using dry natural gas. J Power Sources 123:182–189
14. Lin YB, Zhan ZL, Liu J, Barnett SA (2005) Direct operation of solid oxide fuel cells with methane fuel. Solid State Ionics 176:1827–1835
15. Murray EP, Tsai T, Barnett SA (1999) A direct-methane fuel cell with a ceria-based anode. Nature 400:649
16. Zhan ZL, Barnett SA (2005) An octane-fueled solid oxide fuel cell. Science 308:844–847

17. Bouwmeester HJM, Kruidhof H, Burggraaf AJ (1994) Importance of the surface exchange kinetics as rate limiting step in oxygen permeation through mixed-conducting oxides. Solid State Ionics 72:185
18. van den Bossche M, Matthews R, Lichtenberger A, McIntosh S (2010) Insights into the fuel oxidation mechanism of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$ SOFC anodes. J Electrochem Soc 157:B392–B399
19. van den Bossche M, McIntosh S (2010) Pulse reactor studies to assess the potential of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.4}X_{0.1}O_{3-d}$ (X = co, fe, mn, ni, V) as direct hydrocarbon solid oxide fuel cell anodes. Chem Mater 22:5856–5865
20. van den Bossche M, McIntosh S (2008) Rate and selectivity of methane oxidation over $La_{0.75}Sr_{0.25}Cr_xMn_{1-x}O_{3-d}$ as a function of lattice oxygen stoichiometry under solid oxide fuel cell anode conditions. J Catal 255:313–323
21. McIntosh S, Vohs JM, Gorte RJ (2003) Effect of precious-metal dopants on SOFC anodes for direct utilization of hydrocarbons. Electrochem Solid State Lett 6:A240–A243
22. McIntosh S, Vohs JM, Gorte RJ (2002) An examination of lanthanide additives on the performance of cu-YSZ cermet anodes. Electrochim Acta 47:3815–3821
23. Bouwmeester HJM, Burggraaf AJ (1997) Dense ceramic membranes for oxygen separation. In: Gellings PJ, Bouwmeester HJM (eds) CRC handbook of solid state electrochemistry. CRC Press, Boca Raton
24. Goldschmidt VM (1926) Geochemische verteilungsgesetze der elemente. Akad Oslo I Mat Nat 2:7
25. Knauth P, Tuller HL (2002) Solid-state ionics: roots, status, and future prospects. J Am Ceram Soc 85:1654–1680
26. Oishi M, Yashiro K, Sato K, Mizusaki J, Kawada T (2008) Oxygen nonstoichiometry and defect structure analysis of B-site mixed perovskite-type oxide (la, sr)(cr, M)O_{3-d} (M = ti, mn and fe). J Solid State Chem 181:3177–3184
27. McIntosh S, He HP, Lee SI, Costa-Nunes O, Krishnan VV, Vohs JM, Gorte RJ (2004) An examination of carbonaceous deposits in direct-utilization SOFC anodes. J Electrochem Soc 151: A604–A608
28. McIntosh S, Vohs JM, Gorte RJ (2003) Role of hydrocarbon deposits in the enhanced performance of direct-oxidation SOFCs. J Electrochem Soc 150:A470–A476
29. Coyle CA, Marina OA, Thomsen EC, Edwards DJ, Cramer CD, Coffey GW, Pederson LR (2009) Interactions of nickel/zirconia solid oxide fuel cell anodes with coal gas containing arsenic. J Power Sources 193:730–738
30. Marina OA, Coyle CA, Thomsen EC, Edwards DJ, Coffey GW, Pederson LR (2010) Degradation mechanisms of SOFC anodes in coal gas containing phosphorus. Solid State Ionics 181:430–440
31. Marina OA, Pederson LR, Thomsen EC, Coyle CA, Yoon KJ (2010) Reversible poisoning of nickel/zirconia solid oxide fuel cell anodes by hydrogen chloride in coal gas. J Power Sources 195:7033–7037
32. Gorte RJ, Park S, Vohs JM, Wang C (2000) Anodes for direct oxidation of dry hydrocarbons in a solid-oxide fuel cell. Adv Mater 12:1465–1469
33. Vohs JM, Gorte RJ (2009) High-performance SOFC cathodes prepared by infiltration. Adv Mater 21(9):943–956
34. Kim G, Lee S, Shin JY, Corre G, Irvine JTS, Vohs JM, Gorte RJ (2009) Investigation of the structural and catalytic requirements for high-performance SOFC anodes formed by infiltration of LSCM. Electrochem Solid State Lett 12:B48–B52
35. Lee S, Kim G, Vohs JM, Gorte RJ (2008) SOFC anodes based on infiltration of $La_{0.3}Sr_{0.7}TiO_3$. J Electrochem Soc 155: B1179–B1183
36. Kim H, Park S, Vohs JM, Gorte RJ (2001) Direct oxidation of liquid fuels in a solid oxide fuel cell. J Electrochem Soc 148: A693–A695
37. Mogensen M, Sammes NM, Tompsett GA (2000) Physical, chemical and electrochemical properties of pure and doped ceria. Solid State Ionics 129:63–94
38. Jung S, Lu C, He H, Ahn K, Gorte RJ, Vohs JM (2006) Influence of composition and cu impregnation method on the performance of Cu/CeO_2/YSZ SOFC anodes. J Power Sources 154:42–50
39. Gross MD, Vohs JM, Gorte RJ (2007) An examination of SOFC anode functional layers based on ceria in YSZ. J Electrochem Soc 154:B694–B699
40. Jiang SP, Chan SH (2004) A review of anode materials development in solid oxide fuel cells. J Mater Sci 39:4405–4439
41. Lu C, Worrell WL, Vohs JM, Gorte RJ (2003) A comparison of cu-ceria-SDC and au-ceria-SDC composites for SOFC anodes. J Electrochem Soc 150:A1357–A1359
42. Park S, Gorte RJ, Vohs JM (2001) Tape cast solid-oxide fuel cells for the direct oxidation of hydrocarbons. J Electrochem Soc 148:A443–A447
43. Park S, Vohs JM, Gorte RJ (2000) Direct oxidation of hydrocarbons in a solid-oxide fuel cell. Nature 404:265
44. Kim T, Liu G, Boaro M, Lee S, Vohs JM, Gorte RJ, Al-Madhi OH, Dabbousi BO (2006) A study of carbon formation and prevention in hydrocarbon-fueled SOFC. J Power Sources 155:231–238
45. Kim H, Vohs JM, Gorte RJ (2001) Direct oxidation of sulfur-containing fuels in a solid oxide fuel cell. Chem Commun 22:2334–2335
46. He H, Gorte RJ, Vohs JM (2005) Highly sulfur tolerant cu-ceria anodes for SOFCs. Electrochem Solid State Lett 8:A279–A280
47. Rasmussen JFB, Hagen A (2009) The effect of H_2S on the performance of Ni–YSZ anodes in solid oxide fuel cells. J Power Sources 191:534–541
48. Kim H, Lu C, Worrell WL, Vohs JM, Gorte RJ (2002) Cu-ni cermet anodes for direct oxidation of methane in solid-oxide fuel cells. J Electrochem Soc 149:A247–A250
49. Lee S, Vohs JM, Gorte RJ (2004) A study of SOFC anodes based on cu-ni and cu-co bimetallics in CeO_2-YSZ. J Electrochem Soc 151:A1319–A1323
50. Ahn K, He H, Vohs JM, Gorte RJ (2005) Enhanced thermal stability of SOFC anodes made with CeO_2-ZrO_2 solutions. Electrochem Solid State Lett 8:A414–A417
51. Nakamura T, Petzow G, Gauckler LJ (1979) Stability of the perovskite phase $LaBO_3$ (B = V, cr, mn, fe, co, ni) in reducing atmosphere I. experimental results. Mater Res Bull 14:649

52. Tao SW, Irvine JTS (2003) A redox-stable efficient anode for solid-oxide fuel cells. Nat Mater 2:320–323
53. Lu XC, Zhu JH (2007) Cu(pd)-impregnated $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$ anodes for direct utilization of methane in SOFC. Solid State Ionics 178:1467–1475
54. Doshi R, Alcock CB, Gunasekaran N, Carberry JJ (1993) Carbon-monoxide and methane oxidation properties of oxide solid-solution catalysts. J Catal 140:557–563
55. Sfeir J, Buffat PA, Mockli P, Xanthopoulos N, Vasquez R, Mathieu HJ, Van herle J, Thampi KR (2001) Lanthanum chromite based catalysts for oxidation of methane directly on SOFC anodes. J Catal 202:229–244
56. Danilovic N, Vincent A, Luo J, Chuang KT, Hui R, Sanger AR (2009) Correlation of fuel cell anode electrocatalytic and ex situ catalytic activity of perovskites $La_{0.75}Sr_{0.25}Cr_{0.5}X_{0.5}O_{3-d}$ (X = ti, mn, fe, co). Chem Mater 22(3):957–965
57. Primdahl S, Hansen JR, Grahl-Madsen L, Larsen PH (2001) Sr-doped $LaCrO_3$ anode for solid oxide fuel cells. J Electrochem Soc 148:A74–A81
58. Vernoux P, Djurado E, Guillodo M (2001) Catalytic and electrochemical properties of doped lanthanum chromites as new anode materials for solid oxide fuel cells. J Am Ceram Soc 84:2289–2295
59. Kobsiriphat W, Madsen BD, Wang Y, Marks LD, Barnett SA (2009) $La_{0.8}Sr_{0.2}Cr_{1-x}Ru_xO_{3-d}$-$Gd_{0.1}Ce_{0.9}O_{1.95}$ solid oxide fuel cell anodes: Ru precipitation and electrochemical performance. Solid State Ionics 180:257
60. Tao SW, Irvine JTS (2004) Catalytic properties of the perovskite oxide $La_{0.75}Sr_{0.25}Cr_{0.5}Fe_{0.5}O_{3-d}$ in relation to its potential as a solid oxide fuel cell anode material. Chem Mater 16: 4116–4121
61. Jardiel T, Caldes MT, Moser F, Hamon J, Gauthier G, Joubert O (2010) New SOFC electrode materials: The ni-substituted LSCM-based compounds $(La_{0.75}Sr_{0.25})(Cr_{0.5}Mn_{0.5-x}Ni_x)O_{3-d}$ and $(La_{0.75}Sr_{0.25})(Cr_{0.5-x}Ni_xMn_{0.5})O_{3-d}$. Solid State Ionics 181:894
62. Pudmich G, Boukamp BA, Gonzalez-Cuenca M, Jungen W, Zipprich W, Tietz F (2000) Chromite/titanate based perovskites for application as anodes in solid oxide fuel cells. Solid State Ionics 135:433
63. Hui S, Petric A (2001) Conductivity and stability of $SrVO_3$ and mixed perovskites at low oxygen partial pressures. Solid State Ionics 143:275–283
64. Tao SW, Irvine JTS (2004) Synthesis and characterization of (La0.75Sr0.25)Cr0.5Mn0.5O3-d, a redox-stable, efficient perovskite anode for SOFCs. J Electrochem Soc 151:A252–A259
65. Tao SW, Irvine JTS, Plint SM (2006) Methane oxidation at redox stable fuel cell electrode $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$. J Phys Chem B 110:21771–21776
66. Zha SW, Tsang P, Cheng Z, Liu ML (2005) Electrical properties and sulfur tolerance of $La_{0.75}Sr_{0.25}Cr_{1-x}Mn_xO_3$ under anodic conditions. J Solid State Chem 178:1844–1850
67. Kharton VV, Tsipis EV, Marozau IP, Viskup AP, Frade JR, Irvine JTS (2007) Mixed conductivity and electrochemical behavior of (La0.75Sr0.25)(0.95)Cr0.5Mn0.5O3-delta. Solid State Ionics 178:101–113
68. Plint SM, Connor PA, Tao S, Irvine JTS (2006) Electronic transport in the novel SOFC anode material $La_{1-x}Sr_xCr_{0.5}Mn_{0.5}O_{3+/-d}$. Solid State Ionics 177:2005–2008
69. Wan J, Zhu JH, Goodenough JB (2006) $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$ + Cu composite anode running on H_2 and CH_4 fuels. Solid State Ionics 177:1211–1217
70. Fonseca FC, Muccillo ENS, Muccillo R, de Florio DZ (2008) Synthesis and electrical characterization of the ceramic anode $La_{1-x}Sr_xMn_{0.5}Cr_{0.5}O_3$. J Electrochem Soc 155:B483–B487
71. Kim G, Corre G, Irvine JTS, Vohs JM, Gorte RJ (2008) Engineering composite oxide SOFC anodes for efficient oxidation of methane. Electrochem Solid State Lett 11:B16–B19
72. Jiang SP, Liu L, Khuong POB, Ping WB, Li H, Pu H (2008) Electrical conductivity and performance of doped $LaCrO_3$ perovskite oxides for solid oxide fuel cells. J Power Sources 176:82–89
73. Sfeir J (2003) $LaCrO_3$-based anodes: stability considerations. J Power Sources 118:276–285
74. Liu J, Madsen BD, Ji ZQ, Barnett SA (2002) A fuel-flexible ceramic-based anode for solid oxide fuel cells. Electrochem Solid State Lett 5:A122–A124
75. Yamazoe N, Teraoka Y (1990) Oxidation catalysis of perovskites — relationships to bulk structure and composition (valency, defect, etc.). Catal Today 8:175
76. Pena-Martinez J, Marrero-Lopez D, Ruiz-Morales JC, Savaniu C, Nunez P, Irvine JTS (2006) Anodic performance and intermediate temperature fuel cell testing of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$ at lanthanum gallate electrolytes. Chem Mater 18:1001–1006
77. Tao S, Irvine JTS (2006) Phase transition in perovskite oxide $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_{3-d}$ observed by in situ high-temperature neutron powder diffraction. Chem Mater 18:5453
78. Cheng Z, Zha S, Aguilar L, Liu M (2005) Chemical, electrical, and thermal properties of strontium doped lanthanum vanadate. Solid State Ionics 176:1921–1928
79. Chen XJ, Liu QL, Chan SH, Brandon NP, Khor KA (2007) Sulfur tolerance and hydrocarbon stability of $La_{0.75}Sr_{0.25}Cr_{0.5}Mn_{0.5}O_3$/$Gd_{0.2}Ce_{0.8}O_{1.9}$ composite anode under anodic polarization. J Electrochem Soc 154:B1206–B1210
80. Bernuy-Lopez C, Allix M, Bridges CA, Claridge JB, Rosseinsky MJ (2007) Sr_2MgMoO_{6-d}: structure, phase stability, and cation site order control of reduction. Chem Mater 19:1035–1043
81. Huang YH, Dass RI, Xing ZL, Goodenough JB (2006) Double perovskites as anode materials for solid-oxide fuel cells. Science 312:254–257
82. Otsuka K, Wang Y, Sunada E, Yamanaka I (1998) Direct partial oxidation of methane to synthesis gas by cerium oxide. J Catal 175:152–160
83. Huang YH, Dass RI, Denyszyn JC, Goodenough JB (2006) Synthesis and characterization of Sr_2MgMoO_{6-d} – an anode material for the solid oxide fuel cell. J Electrochem Soc 153: A1266–A1272
84. Marrero-López D, Peña-Martínez J, Ruiz-Morales JC, Gabás M, Núñez P, Aranda MAG, Ramos-Barrado JR (2009) Redox behaviour, chemical compatibility and electrochemical performance of Sr_2MgMoO_6 – [delta] as SOFC anode. Solid State Ionics 180:1672

85. Vasala S, Lehtimaki M, Haw SC, Chen JM, Liu RS, Yamauchi H, Karppinen M (2010) Isovalent and aliovalent substitution effects on redox chemistry of Sr_2MgMoO_{6-d} SOFC-anode material. Solid State Ionics 181:754–759
86. Matsuda Y, Karppinen M, Yamazaki Y, Yamauchi H (2009) Oxygen-vacancy concentration in A_2MgMoO_{6-d} double-perovskite oxides. J Solid State Chem 182:1713–1716
87. Huang Y, Liang G, Croft M, Lehtimaki M, Karppinen M, Goodenough JB (2009) Double-perovskite anode materials Sr_2MMoO_6 (M = co, ni) for solid oxide fuel cells. Chem Mater 21:2319–2326
88. Vasala S, Lehtimäki M, Huang YH, Yamauchi H, Goodenough JB, Karppinen M (2010) Degree of order and redox balance in B-site ordered double-perovskite oxides, Sr_2MMoO_{6-d} (M = Mg, mn, fe, co, ni, zn). J Solid State Chem 183:1007
89. Marrero-Lopez D, Pena-Martinez J, Ruiz-Morales JC, Martin-Sedeno MC, Nunez P (2009) High temperature phase transition in SOFC anodes based on Sr_2MgMoO_6-delta. J Solid State Chem 182:1027–1034
90. Marrero-Lopez D, Pena-Martinez J, Ruiz-Morales JC, Perez-Coll D, Aranda MAG, Nunez P (2008) Synthesis, phase stability and electrical conductivity of Sr_2MgMoO_{6-d} anode. Mater Res Bull 43:2441–2450
91. Moos R, Hardtl KH (1997) Defect chemistry of donor-doped and undoped strontium titanate ceramics between 1000° and 1400 °C. J Am Ceram Soc 80:2549–2562
92. Hui S, Petric A (2002) Evaluation of yttrium-doped $SrTiO_3$ as an anode for solid oxide fuel cells. J Eur Ceram Soc 22:1673–1681
93. Kolodiazhnyi T, Petric A (2005) The applicability of sr-deficient n-type SrTiO3 for SOFC anodes. J Electroceram 15:5–11
94. Lu XC, Zhu JH, Yang Z, Xia G, Stevenson JW (2009) Pd-impregnated SYT/LDC composite as sulfur-tolerant anode for solid oxide fuel cells. J Power Sources 192: 381–384
95. Moos R, Bischoff T, Menesklou W, Hardtl K (1997) Solubility of lanthanum in strontium titanate in oxygen-rich atmospheres. J Mater Sci 32:4247–4252
96. Marina OA, Canfield NL, Stevenson JW (2002) Thermal, electrical, and electrocatalytical properties of lanthanum-doped strontium titanate. Solid State Ionics 149:21–28
97. Huang X, Zhao H, Shen W, Qiu W, Wu W (2006) Effect of fabrication parameters on the electrical conductivity of YxSr1 − xTiO3 for anode materials. J Phys Chem Solids 67:2609–2613
98. Balachandran U, Eror NG (1982) Electrical conductivity in lanthanum-doped strontium titanate. J Electrochem Soc 129:1021–1026
99. Flandermeyer BF, Agarwal AK, Anderson HU, Nasrallah MM (1984) Oxidation-reduction behaviour of la-doped $SrTiO_3$. J Mater Sci 19:2593–2598
100. Battle PD, Bennett JE, Sloan J, Tilley RJD, Vente JF (2000) A-site cation-vacancy ordering in Sr1 − 3x/2LaxTiO3: A study by HRTEM. J Solid State Chem 149:360–369
101. Meyer R, Waser R, Helmbold J, Borchardt G (2002) Cationic surface segregation in donor-doped $SrTiO_3$ under oxidizing conditions. J Electroceram 9:101–110
102. Fu QX, Mi SB, Wessel E, Tietz F (2008) Influence of sintering conditions on microstructure and electrical conductivity of yttrium-substituted SrTiO3. J Eur Ceram Soc 28:811–820
103. Slater PR, Fagg DP, Irvine JTS (1997) Synthesis and electrical characterisation of doped perovskite titanates as potentialanode materials for solid oxide fuel cells. J Mater Chem 7:2495–2498
104. Blennow P, Hagen A, Hansen KK, Wallenberg LR, Mogensen M (2008) Defect and electrical transport properties of nb-doped $SrTiO_3$. Solid State Ionics 179:2047–2058
105. Blennow P, Hansen KK, Wallenberg LR, Mogensen M (2009) Electrochemical characterization and redox behavior of nb-doped $SrTiO_3$. Solid State Ionics 180:63–70
106. Blennow P, Hansen KK, Wallenberg LR, Mogensen M (2007) Synthesis of nb-doped SrTiO3 by a modified glycine-nitrate process. J Eur Ceram Soc 27:3609–3612
107. Hashimoto S, Poulsen FW, Mogensen M (2007) Conductivity of SrTiO3 based oxides in the reducing atmosphere at high temperature. J Alloy Compd 439:232–236
108. Tufte ON, Chapman PW (1967) Electron mobility in semiconducting strontium titanate. Phys Rev 155:796
109. Irvine J, Slater P, Wright P (1996) Synthesis and electrical characterisation of the perovskite niobate-titanates, Sr1 − x/2Ti1 − xNbxO3 − δ. Ionics 2:213–216
110. Blennow P, Hansen KK, Reine Wallenberg L, Mogensen M (2006) Effects of Sr/Ti-ratio in SrTiO3-based SOFC anodes investigated by the use of cone-shaped electrodes. Electrochim Acta 52:1651–1661
111. Hui S, Petric A (2002) Electrical conductivity of yttrium-doped $SrTiO_3$: Influence of transition metal additives. Mater Res Bull 37:1215–1231
112. Zhao H, Gao F, Li X, Zhang C, Zhao Y (2009) Electrical properties of yttrium doped strontium titanate with A-site deficiency as potential anode materials for solid oxide fuel cells. Solid State Ionics 180:193
113. Fu Q, Tietz F, Sebold D, Tao S, Irvine JTS (2007) An efficient ceramic-based anode for solid oxide fuel cells. J Power Sources 171:663–669
114. Yang LM, De Jonghe LC, Jacobsen CP, Visco SJ (2007) B-site doping and catalytic activity of sr(Y)TiO_3. J Electrochem Soc 154:B949–B955
115. Vincent A, Luo J, Chuang KT, Sanger AR (2010) Effect of ba doping on performance of LST as anode in solid oxide fuel cells. J Power Sources 195:769
116. Wagner SF, Warnke C, Menesklou W, Argirusis C, Damjanovic T, Borchardt G, Ivers-Tiffée E (2006) Enhancement of oxygen surface kinetics of $SrTiO_3$ by alkaline earth metal oxides. Solid State Ionics 177:1607
117. Sun X, Wang S, Wang Z, Qian J, Wen T, Huang F (2009) Evaluation of $Sr_{0.88}Y_{0.08}TiO_3$–CeO_2 as composite anode for solid oxide fuel cells running on CH4 fuel. J Power Sources 187:85–89

118. Sun X, Wang S, Wang Z, Ye X, Wen T, Huang F (2008) Anode performance of LST–xCeO2 for solid oxide fuel cells. J Power Sources 183:114–117
119. Fu QX, Tietz F, Stover D (2006) $La_{0.4}Sr_{0.6}Ti_{1-x}Mn_xO_{3-d}$ perovskites as anode materials for solid oxide fuel cells. J Electrochem Soc 153:D74–D83
120. Ruiz-Morales J, Canales-Vázquez J, Savaniu C, Marrero-López D, Zhou W, Irvine JTS (2006) Disruption of extended defects in solid oxide fuel cell anodes for methane oxidation. Nature 439:568–571
121. Cheng Z, Zha S, Liu M (2006) Stability of materials as candidates for sulfur-resistant anodes of solid oxide fuel cells. J Electrochem Soc 153:A1302–A1309
122. Mukundan R, Brosha EL, Garzon FH (2004) Sulfur tolerant anodes for SOFCs. Electrochem Solid State Lett 7:A5–A7
123. Zha SW, Cheng Z, Liu ML (2005) A sulfur-tolerant anode material for SOFCs $Gd_2Ti_{1.4}Mo_{0.6}O_7$. Electrochem Solid State Lett 8:A406–A408
124. Porat O, Heremans C, Tuller HL (1997) Stability and mixed ionic electronic conduction in $Gd_2(Ti_{1-x}Mo_x)_2O_7$ under anodic conditions. Solid State Ionics 94:75
125. Sprague JJ, Tuller HL (1999) Mixed ionic and electronic conduction in Mn/Mo doped gadolinium titanate. J Eur Ceram Soc 19:803–806
126. Deng ZQ, Niu HJ, Kuang XJ, Allix M, Claridge JB, Rosseinsky MJ (2008) Highly conducting redox stable pyrochlore oxides. Chem Mater 20:6911–6916
127. Kramer S, Spears M, Tuller HL (1994) Conduction in titanate pyrochlores – role of dopants. Solid State Ionics 72:59–66
128. Naito H, Arashi H (1992) Electrical-properties of ZrO_2-TiO_2-Y_2O_3 system. Solid State Ionics 53:436–441
129. Kaiser A, Feighery A, Fagg D, Irvine J (1998) Electrical characterization of highly titania doped YSZ. Ionics 4:215
130. Tao S, Irvine JTS (2002) Optimization of mixed conducting properties of Y_2O_3-ZrO_2-TiO_2 and Sc_2O_3-Y_2O_3-ZrO_2-TiO_2 solid solutions as potential SOFC anode materials. J Solid State Chem 165:12–18
131. Tao SW, Irvine JTS (2004) Investigation of the mixed conducting oxide ScYZT as a potential SOFC anode material. J Electrochem Soc 151:A497–A503
132. Slater PR, Irvine JTS (1999) Niobium based tetragonal tungsten bronzes as potential anodes for solid oxide fuel cells: synthesis and electrical characterisation. Solid State Ionics 120:125–134
133. Slater PR, Irvine JTS (1999) Synthesis and electrical characterisation of the tetragonal tungsten bronze type phases, $(Ba/Sr/Ca/La)_{0-6}M_xNb_{1-x}O_{3-d}$ (M = mg, ni, mn, cr, fe, in, sn): evaluation as potential anode materials for solid oxide fuel cells. Solid State Ionics 124:61–72
134. Kaiser A, Bradley JL, Slater PR, Irvine JTS (2000) Tetragonal tungsten bronze type phases $(Sr_{1-x}Ba_x)_{0.6}Ti_{0.2}Nb_{0.8}O_{3-d}$: material characterisation and performance as SOFC anodes. Solid State Ionics 135:519–524
135. Aguilar L, Zha S, Cheng Z, Winnick J, Liu M (2004) A solid oxide fuel cell operating on hydrogen sulfide (H2S) and sulfur-containing fuels. J Power Sources 135:17–24

Books and Reviews

Atkinson A, Barnett S, Gorte RJ, Irvine JTS, McEvoy AJ, Mogensen M, Singhal SC, Vohs J (2004) Advanced anodes for high-temperature fuel cells. Nat Mater 3:17–27
Brandon NP, Skinner S, Steele BCH (2003) Recent advances in materials for fuel cells. Annu Rev Mater Res 33:183–213
Fergus JW (2006) Oxide anode materials for solid oxide fuel cells. Solid State Ionics 177:1529
Fleig J (2003) Solid oxide fuel cell cathodes: polarization mechanisms and modeling of the electrochemical performance. Annu Rev Mater Res 33:361–382
Gellings PJ, Bouwmeester HJM (2000) Solid state aspects of oxidation catalysis. Catal Today 58:1–53
Goodenough JB (2003) Oxide-ion electrolytes. Annu Rev Mater Res 33:91–128
Goodenough JB, Huang Y (2007) Alternative anode materials for solid oxide fuel cells. J Power Sources 173:1
Gong M, Liu X, Trembly J, Johnson C (2007) Sulfur-tolerant anode materials for solid oxide fuel cell application. J Power Sources 168:289
Gorte RJ, Vohs JM (2009) Nanostructured anodes for solid oxide fuel cells. Curr Opin Colloid Interface Sci 14:236–244

Disease-Resistant Transgenic Animals

Caroline Lassnig, Mathias Müller
Institute of Animal Breeding and Genetics & Biomodels Austria, University of Veterinary Medicine Vienna, Vienna, Austria

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Disease-Resistant Transgenic Animals
Future Directions
Bibliography

Glossary

Disease resistance/susceptibility The interplay of the genetically determined ability of an individual to prevent the reproduction of a pathogen or to reduce pathogen growth, host–pathogen interactions, and changing environmental conditions/factors decides on resistance/susceptibility.

Gene targeting Integration of exogenous DNA into the genome of an organism at specific sites as a result of homologous recombination. It can be used to disrupt or delete a gene, to remove or add sequences as well as to introduce point mutations at a given locus. Gene targeting can be permanent, i.e., ubiquitous with respect to tissue and developmental stage, or conditional, i.e., restricted to a specific time during development/life or limited to a specific tissue.

Genetic engineering Technological process resulting in a directed alteration of the genotype of a cell or organism. It combines recombinant nucleic acid technologies, in vitro culture technologies for gametes, embryos, tissues, or organisms, methods for the delivery of nucleic acids to the host genomes (gene transfer), and if needed, reproductive technologies to produce transgenic embryos and transfer them to foster organisms. With respect to inheritance ('transmission') to offspring, germline and somatic gene transfer methods are distinguished.

Knockdown Downregulation of expression of a specific gene by RNAi-based technologies.

Knockout/knockin Incorporation of a sequence into a specific site by homologous recombination (gene targeting) that results in disruption of gene function/altered gene function.

Quantitative trait loci (QTL) Genetic loci or chromosomal regions that contribute to variability in complex traits, as identified by statistical analysis. The genetic basis of these traits generally involves the effects of multiple genes and gene–environment interactions.

RNA interference (RNAi) The silencing of gene expression by the introduction of dsRNA that triggers the specific degradation of a homologous target mRNA, often accompanied by an attendant decrease in the production of the encoded protein.

Single nucleotide polymorphism (SNP) A variation in DNA sequence in which one nucleotide position is substituted for another by either nucleotide exchange, or deletion, or insertion. SNPs are the most frequent type of polymorphism in the genome.

Somatic cell nuclear transfer (SCNT) The nonsexual generation of nuclear genome-identical offspring ("cloned animals") by reconstitution of an enucleated oocyte with the diploid nucleus of a somatic cell to a zygote, which under appropriate culture conditions leads to reprogramming of the genome, enabling embryonic and fetal development.

Zoonotic infection The ability of a given pathogen to cross the host species barrier, from its current or long-term evolutionary host to animals and humans and thereby causing disease.

Definition of the Subject and Its Importance

Infectious diseases of livestock are a major risk to global animal health and welfare. In addition, human health is influenced due to the zoonotic potential of some of these infections.

Moreover, livestock diseases significantly impair food production and safety and cause enormous economic losses worldwide.

Transgenic technology was first developed as a research tool for studying gene function in mice in the early 1980s. The technique was extended and applied to other mammals in 1985. An interesting and challenging focus of agricultural transgenesis was the potential to increase disease resistance and/or reduce disease susceptibility by introducing new genes and/or deleting deleterious genes. The laborious improvements to original and recently developed transgene technologies lead to the generation of transgenic farm animals with improved resistance to infectious diseases, demonstrating the proof of principle that genetic engineering may potentially improve animal health and aid infectious disease control in livestock.

Introduction

Phenotype-driven traditional animal breeding and marker-assisted selection based on quantitative trait loci (QTLs) have been successfully used for the genetic improvement of many agricultural production traits such as body weight, carcass composition, or milk yield. However, these genetic selection strategies have not yet resulted in a significant increase in the resistance of farm animals to disease.

Currently, genomic sequences are available for several livestock species [1] and as a by-product of the sequencing, a huge number of single nucleotide

polymorphisms (SNPs) were discovered. The large panels of available SNPs were used in genome-wide association (GWA) studies for mapping and identifying genes [2]. GWA studies have already been successful in identifying causal genes and mutations for monogenic traits [3], but not for complex or quantitative traits such as resistance or susceptibility to disease.

Furthermore, traditional strategies in combating devastating infectious diseases of livestock, such as vaccination, antibiotic treatment, or even culling, have, to date, been unsuccessful (Fig. 1). Parasites evolved to resist chemical or vaccine control measures and bacteria developed resistance to many antibiotics. So far, a single infectious viral disease in livestock, rinderpest (cattle plague) could be eliminated through large-scale vaccination.

As an alternative to the traditional approaches, genetic engineering of livestock species may assist in the fight against infectious diseases.

The oldest and probably the most robust technique to produce transgenic farm animals is the injection of DNA sequences into the pronucleus of recently fertilized zygotes [4–6]. Pronuclear microinjection was successfully used to generate the most important livestock species, mainly for production of highly valuable human therapeutics. A more recent method for generating transgenic animals is the nuclear transfer technology, that is, 'cloning' [7, 8], which, together with a gene-targeting strategy allows the generation of specific gene-targeted animals [9, 10]. Recently, lentiviral vector-based strategies have been established which results in highly efficient production of transgenic livestock [11, 12]. This method in combination with the RNAi-technology may lead to the generation of disease-resistant transgenic livestock in the near future [13].

In the following section, the authors present an overview of the various transgenic methods used for the genetic enhancement of animal resistance to infectious diseases. Many studies were initially done using transgenic mouse models as this model often provides useful preliminary results prior to initiation of livestock studies.

Disease-Resistant Transgenic Animals

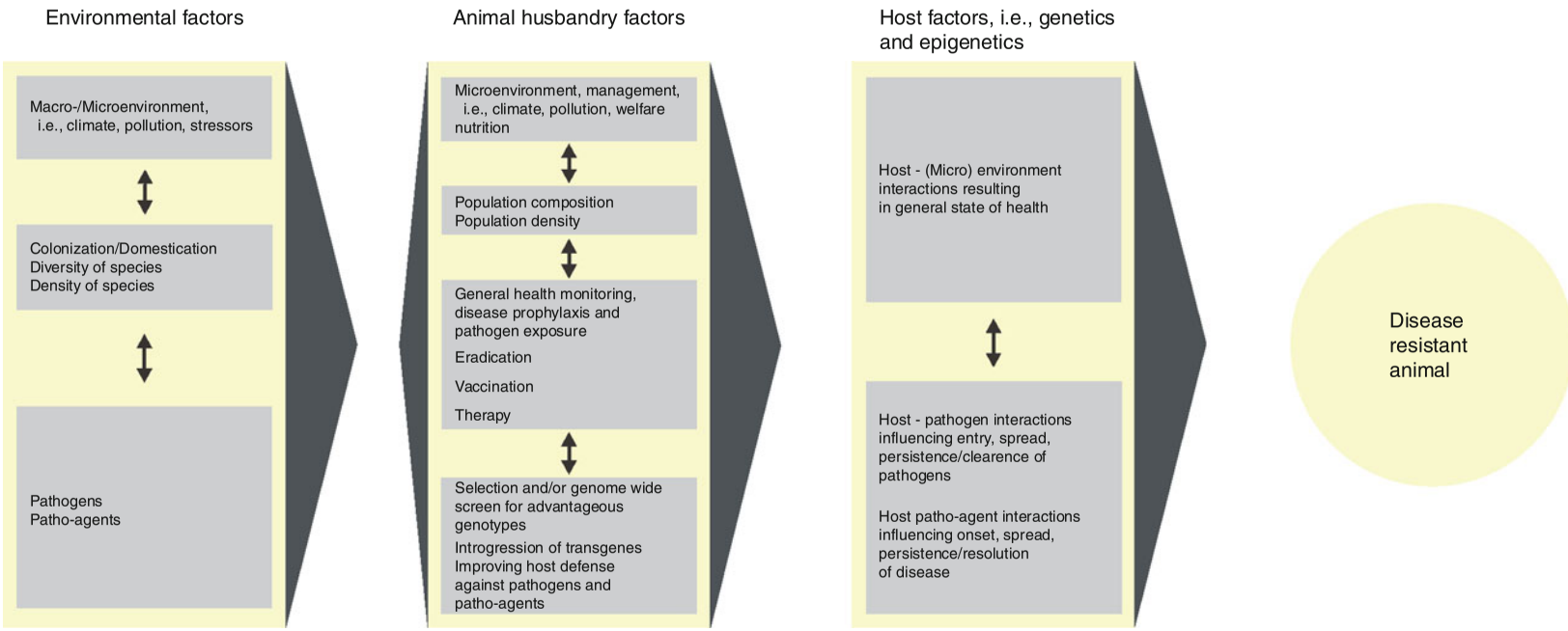
Reducing farm animal susceptibilty to infectious diseases via genetic engineering has been an ambitious goal since the first transgenic livestock was generated more than 20 years ago. Various transgenic strategies for improving animal health are described elsewhere [14–17].

In general, disease-resistant transgenic farm animals can be generated by two approaches: (1) introduction of resistance genes into the genome of the host (gain-of-function strategy) and (2) specific targeting of endogenous or exogenous susceptibility genes (loss or exchange-of-function strategy).

Improving Animal Health through Gain-of-Function Gene Transfer

In most cases, susceptibility to pathogens originates from the interplay of numerous genes, meaning susceptibility to pathogens is polygenic in nature. The murine *Mx* gene is one of the few examples of a single genetic (monogenic) locus encoding a disease-resistance trait. Mice and mouse fibroblast cell lines carrying the autosomal dominant *Mx1* allele are resistant to influenza virus infection [18, 19]. The transfer of the *Mx1* gene was able to restore virus resistance in mice lacking the *Mx1* allele [20] and inhibited influenza virus replication in avian cells [21]. However, the introduction of the murine *Mx1* gene into swine via pronuclear microinjection failed to produce influenza-resistant pigs [22]. The constitutive *Mx1* expression seemed to be detrimental to the pigs, whereas the expression from an inducible promoter was too low to produce detectable levels of *Mx1* protein. In the meantime, *Mx* genes of different farm animals have been identified, but their importance for disease susceptibility is not yet clear [23–25]. However, the ongoing detailed deciphering of the genomes of different farm animals, the improved techniques in generating transgenic animals [26–28], and the new tools for controlling transgene expression levels [29, 30] might allow the idea of generating influenza-resistant livestock by transferring a disease resistance gene to be addressed once more.

Antimicrobial peptides (AMPs) are an important component of the innate defense of most living organisms and there is a growing body of evidence to show that their role in defense against microbes is as important to the host as antibodies and innate and adaptive immune cells [31, 32]. AMPs are usually composed of



Disease-Resistant Transgenic Animals. Figure 1

Horizontal and vertical interrelations influencing the animal's health status, diseased versus nondiseased. The most specific measures to disrupt the pathogen/patho-agent flow toward the animal and/or improve the animal's disease defense mechanisms are highlighted

12–50 amino acids and synthesized by microorganisms as well as multicellular organisms, including plants and animals. They can have broad-spectrum antibacterial, antifungal, antiviral, antiprotozoan, and antisepsis properties. In addition to the wide range of these naturally occurring AMPs, many new ones have also been synthesized [33, 34]. Based on three-dimensional structural studies, the peptides are broadly classified into five major groups namely: (1) peptides that form alpha-helical structures; (2) peptides that form beta-sheets; (3) peptides rich in cysteine residues; (4) peptides rich in regular amino acids namely histatin, arginine, and proline; and (5) peptides composed of rare and modified amino acids [35, 36]. They can induce complete lysis of the organism by disrupting the membrane or by perturbing the membrane lipid bilayer, which allows for leakage of specific cellular components as well as dissipating the electrical potential of the membrane.

In initial engineering studies, the endogenous production of antimicrobial compounds in transgenic animals was shown to enhance disease resistance. Recombinant bovine tracheal antimicrobial peptide (bTAP) isolated from milk from transgenic mice, showed antimicrobial activity against *Escherichia coli*, without any deleterious side effects in suckling pups [37]. The antimicrobial activity of the synthetic alpha-helical peptide *Shiva 1a* was confirmed in transgenic mice, challenged with *Brucella abortus* [38]. The expression of the recombinant peptide significantly reduced both the bacterial colonization and the associated pathological changes in the genetically engineered mice.

Mastitis which is caused by bacterial infection of the mammary gland is reported to be the most costly disease in animal agriculture. It seriously affects animal well-being and is the most common reason for antibiotic use in diary cattle and the most frequent cause of antibiotic residues in milk [39]. The major contagious mastitis pathogen, *Staphylococcus aureus* is sensitive to lysostaphin, an antibacterial peptide naturally produced by a related bacterium, *Staphylococcus simulans* [40]. Kerr and colleagues showed that mammary gland expression of a bioactive variant of lysostaphin conferred protection against *S.aureus* infection in mice [41]. The staphylolytic activity in the milk of transgenic mice appeared to be 5–10 fold less active than bacterially derived lysostaphin, but was sufficient to confer substantial resistance to staphylococcal mastitis. Transgene production appeared to have no apparent effect on the physiology of the animal, the integrity of the mammary gland, or the milk it produces. Using nuclear transfer techniques, this approach was successfully extended to cattle, recently [42]. Transgenic dairy cows secreting lysostaphin constitutively in their milk were more resistant to *S. aureus* infections than nontransgenic animals. Lysostaphin concentrations in the milk of transgenic animals remained fairly constant during lactation. The recombinant lysostaphin was approximately 15% as active as bacterially derived protein. Challenge studies with *S. aureus* clearly demonstrated a direct correlation between the extent of protection against *S. aureus* infection with lysostaphin levels in the milk. Transgenic cows have been previously generated, primarily as bioreactors for large-scale production of pharmaceuticals and nutraceuticals. Thus, lysostaphin-transgenic cattle are the first example for enhancing disease resistance and animal welfare in livestock, and may allow substantial reductions in antibiotic use. This in turn will help to control the spread of antibiotic-resistant bacteria and to reduce bacterial and antibiotic contamination of milk and milk products.

The antibacterial effect of lysostaphin is restricted to *S. aureus* only and transgenic cows are not protected against other mastitis-causing pathogens. The additional expression of secondary antibacterial compounds in the milk might be necessary for further enhancing mastitis resistance.

Human lysozyme (hLZ), a bacteriostatic milk protein that is known to attack the peptidoglycan component of bacterial cell walls, was expressed in the mammary gland of transgenic mice [43] and transgenic dairy goats [44]. Milk from the transgenic animals showed significant bacteriostatic activity and slowed the growth of several bacteria responsible for causing mastitis and the cold-spoilage of milk. The somatic cell count (SCC) is applied as a measure for udder health and milk quality and a high SCC in milk is directly correlated with mastitis and an impairment of milk quality [45]. Analyzing the SCC in milk samples of transgenic diary goats revealed a significant lower SCC compared to milk samples from control animals suggesting an improved udder health in the transgenic animals [46]. Lysozyme plays a role in the defense

against gastrointestinal pathogens and reduces gastrointestinal illness in breastfed infants [47]. Feeding trials were conducted in pigs to evaluate putative health-promoting functions of hLZ-transgenic milk. Pigs are monogastric animals with a digestive system similar to humans and therefore are commonly used to study human health. Brundige and colleagues demonstrated that the consumption of pasteurized milk from hLZ-transgenic goats improved the gastrointestinal health of young piglets and was beneficial against a gastrointestinal infection with enteropathogenic *E. coli* [48].

A Chinese group enabled synthesis and secretion of bioactive bovine lactoferrin and bovine tracheal antibacterial peptides in goat mammary cells by use of plasmid-mediated gene transfer techniques [49], and the milk samples collected from these animals exhibited bacteriostatic effects against different mastitis-causing pathogens.

The authors summarize that genetic engineering for secretion of a broad range of AMPs in the mammary gland of dairy goats and cows reduces susceptibility to various microbial pathogens and is therefore a realistic approach to combat mastitis. Enhanced mastitis resistance will not only improve animal health and well-being, but also reduces bacterial contamination of milk and milk products in addition to reducing the costs incurred during disease prevention and cure.

Transgenic mice, expressing and processing a human enteric alpha-defensin peptide exclusively in specialized epithelia of the small intestinal crypt were generated, and were immune to an oral challenge with virulent *Salmonella typhimurium* [50].

Protegin-1 (PG-1) that is normally expressed in porcine myeloid cells and resides in secretory granules of neutrophils is another potent antimicrobial peptide targeting both gram-negative and gram-positive bacteria [51]. The ectopic expression of PG-1 in transgenic mice conferred enhanced respiratory resistance to an intranasal challenge with *Actinobacillus suis* [52], an opportunistic pathogen that may cause pneumonia, abortion, and fatal septicemia in pigs of all ages [53, 54]. Extending this concept to pigs and other somatic tissues beyond neutrophils will be another step toward the development of disease-resistant livestock.

The overexpression of dominant-negative mutants of viral proteins or pathogen receptors is another potent strategy to enhance animal disease resistance. The major focus has been to block viral attachment and penetration into a host cell by (1) producing viral proteins that block cellular receptors (antireceptor) or (2) altering known host molecular components, such as replacing host receptor genes with a modified version which is able to perform the receptor's physiological function but prevents attachment of the virus [55]. The first successful introduction of pathogen-mediated disease resistance in animals was reported 20 years ago. Transgenic chickens expressing the viral envelope of a recombinant avian leukosis retroviral genome were resistant to the corresponding subgroup of avian leukosis virus due to blockage of the virus receptors by the viral envelope proteins [56]. Using the same strategy, Clements et al. generated transgenic sheep expressing the maedi-visna virus envelope (E) gene, which is responsible for virus attachment to the host cells [57]. Maedi-visna virus is a prototype of ovine lentiviruses that cause encephalitis, pneumonia, and arthritis in sheep. Transgenic lambs expressing the viral E glycoprotein in monocytes/macrophages, the target cells for virus replication, were healthy and neither deleterious effects nor clinical abnormalities from the transgene was observed. However, up to date, challenge studies to determine the susceptibility of these animals to ovine lentiviruses have not been reported.

Transgenic mice expressing a soluble form of porcine nectin-1, the cellular receptor for α-herpesviruses were generated. These mice displayed high resistance to pseudorabies virus (PRV) infections [58]. In pigs, PRV causes lethal encephalitis, acute respiratory syndrome, abortion and infertility, and latent infections [59]. Analysis of transgenic mouse lines, ubiquitously expressing different soluble forms of the cellular receptor for the viral glycoprotein D revealed that the transgene encoding the soluble form of the entire ectodomain of porcine nectin-1 fused to the human IgG1 conferred highest resistance to intranasal and intraperitonal PRV infections without any side effects [60]. Surprisingly, the expression of a fusion protein consisting of the first Ig-like domain of nectin-1 and the Fc portion of porcine IgG1 not only resulted in reduced virus resistance but also caused microphthalmia and the lack of vitreous bodies [61, 62]. Before implementing this promising approach to the generation of α-herpesviruses-resistant swine, further investigations examining the

interactions of different soluble forms of nectin-1, endogenous nectins, and viral glycoprotein D and analysis of the influence of Fc domains of different species are required.

An alternative transgenic approach to protect livestock against infectious diseases is the expression of genes directing the synthesis of defined antibodies which target specific pathogens and thus induce immediate immunity without prior exposure to that pathogen.

Initial studies to express gene constructs encoding monoclonal antibodies in transgenic livestock were conducted nearly 20 years ago [63, 64]. However, the recombinant antibodies expressed in transgenic rabbits, sheep, and pigs showed aberrant sizes and only low antigen binding affinity. Nevertheless, following this idea, transgenic mice expressing coronavirus-neutralizing antibodies in the mammary gland were generated [65, 66]. High antibody expression titres throughout the lactation period provided complete protection against the enteric infection of newborns with transmissible gastroenteritis virus (TGEV), a pathogen which produces high mortality in suckling piglets, and also against a murine hepatitis virus (MHV)-induced encephalitis. Following this strategy, manipulating the lactogenic immunity in farm animals could improve the protection of suckling newborns through colostrum-delivered antibodies [67].

Enhancing Disease Resistance by Targeting Endogenous Susceptibility Genes

Transmissible spongiform encephalopathies (TSE) are fatal neurodegenerative disorders of the central nervous system which are termed scrapie in goat and sheep and bovine spongiform encephalopathy (BSE) in cattle. According to current knowledge, the causative agent of the brain pathology in diseased animals is the prion. Prion diseases are characterized by the accumulation of the abnormally folded and protease-resistant isoform (PrP^{Sc}) of the cellular prion protein (PrP^{C}) of the host [68, 69]. The generation of prion-free livestock resistant to TSE has been an ambitious goal since the BSE epidemic in cattle in the UK and the appearance of a new and highly lethal variant of Creutzfeldt-Jakob disease (vCJD) in humans. Early studies in mice revealed that reduction or loss of PrP^{C} expression did not affect normal development of the mice, but conferred protection against scrapie disease after inoculation with PrP^{Sc} prions [70–73]. With the development of nuclear transfer cloning techniques using genetically modified embryonic or somatic cell donors [7–10], the possible 'knock out' of the prion gene in transgenic sheep, goats, and cattle has opened new perspectives for the generation of disease-resistant livestock. A decade ago, Denning and colleagues generated the first PrP^{C}-targeted lambs. However, none of the cloned sheep survived more than 12 days [74]. Analyses of the targeted fetuses and lambs revealed defects that have been described in other nuclear transfer experiments with nontransfected cells and therefore, the authors expected that the early death of the lambs was not a consequence of the PrP^{C} disruption per se, but was probably due to the nuclear transfer procedures and/or the prolonged culture and drug selection of the primary fibroblasts used for nuclear transfer.

The functional disruption of the caprine PrP^{C} gene in cloned goats was first described by Yu et al. [75] and resulted in two goats lacking the prion protein [76]. The scientists confirmed the complete PrP^{C} ablation at mRNA and protein levels, and at 2 months age, the PrP^{C} null goats were healthy and showed no developmental or behavioral defects. The scientific community is awaiting the final proof of the concept – scrapie resistance of PrP^{C}-deficient goats after infection with PrP^{Sc} prions.

Richt and colleagues described the generation of the first PrP^{C}-deficient cattle [77]. They used a sequential gene targeting strategy which was demonstrated for the first time by the group of Kuroiwa et al. [78]. Male Holstein primary fibroblasts were transfected with two knockout vectors to sequentially disrupt the two alleles of the PrP^{C} gene. PrP^{C}-deficient fetal cell lines were established at 40–75 days of gestation and recloned for the generation of calves. The impact of PrP^{C} deficiency on calf development, on the immune system, on growth, and general health of the cattle for at least 20 months was analyzed in detail, and no negative influence of PrP^{C} ablation on animal health and well-being was detected. Importantly, brain homogenates from 10-month-old PrP^{C}-deficient cattle prevented PrP^{Sc} propagation in vitro, whereas in brain homogenates from wild-type cattle PrP^{Sc} proliferated. The researchers concluded that the presence of the

endogenous bovine PrP^C is essential for PrP^{Sc} propagation and that there are no other host-derived cellular factors that can support the in vitro PrP^{Sc} propagation in the absence of the endogenous bovine PrP^C. In vivo tests of resistance to prion propagation in PrP^C-deficient cattle are under way, but still will require some years to complete. Analyses of several PrP^C-targeted mouse lines indicated that the loss of the normal cellular function of PrP^C may adversely affect the animals. For example, PrP^C-deficient mice developed ataxia and cerebellar neurodegeneration [79, 80], slight alterations in sleep–wake circadian rhythm [81], and altered synaptic functions [82]. To date, none of the above-described alterations in PrP^C null mice could be observed in PrP^C-deficient cattle and goats, respectively, but further investigations on aged transgenic animals will be necessary to exclude these altered phenotypes.

Small interfering RNAs (siRNAs) can silence/shut down specific targeted genes by interfering with the RNA transcripts they produce [83, 84]. For a transient gene 'knock down,' synthetic siRNAs can be directly transfected into cells or early embryos. However, for stable gene expression and germline transmission, the siRNA sequences are incorporated into gene constructs which express short hairpin (sh)RNAs that are processed to siRNAs within the cell. Through stably integrated shRNA expression vectors, additional genetic information is introduced into an organism (gain-of-functions strategy), which then produces a 'knock down' phenotype that is functionally similar to a 'knock out' (loss of function). Thus, RNAi-transgenics is an interesting alternative to the homologous gene targeting strategies which are traditionally used for the generation of 'knock out' livestock.

One of the most interesting susceptibility genes in livestock is the PrP^C gene and in a preliminary in vitro experiment, it was demonstrated that siRNA suppression of the PrP^C gene abrogates the PrP^C synthesis and inhibits the formation of PrP^{SC} protein in chronically scrapie-infected murine neuroblastoma cells [85]. Shortly after, Golding and colleagues combined this RNAi-based technique with lentiviral transgenesis for targeting the PrP^C gene in an adult goat fibroblast cell line, which was then used for somatic cell nuclear transfer to produce a cloned goat fetus [13]. Protein analyses of brain tissues demonstrated that PrP^C expression was reduced >90% in the cloned transgenic fetus when compared with a control. In a further experiment, they injected the recombinant lentivirus directly into the perivitelline space of bovine ova. Development of more than 30% of injected ova to blastocysts and expression of the shRNA targeting the PrP^C gene in more than 70% provides strong evidence that this RNAi approach may be useful in creating genetically engineered farm animals with natural resistance to prion diseases.

In two further approaches, lentiviral-mediated delivery of shRNA expression vectors into the brain of scrapie-infected mice resulted in a clear reduction of the PrP^C protein level and a prolonged survival of infected mice [86, 87], inferring that RNAi-technology may also be used for therapeutic applications.

Enhancing Disease Resistance by Targeting Exogenous Susceptibility/Viral Genes

Another application of the RNAi-technology is the silencing of exogenous viral genes through the introduction of specific dsRNA molecules into cells, where they are targeted to essential genes or directly to the viral genome, thus inhibiting viral replication [88, 89]. Currently, the use of RNAi-based strategies for generation of viral disease-resistant livestock focuses on three pathogens: food and mouth disease virus (FMDV), bovine viral diarrhea virus (BVDV), and influenza A viruses.

FMDV is an extremely contagious pathogen that affects cattle, swine, and other livestock worldwide [90]. FMD is difficult to control by vaccination and impossible to eliminate by conservative natural breeding. Initial studies tested specific FMDV-siRNAs for their ability to inhibit virus replication in BHK-21 cells [91]. Transfection of BHK-21 cells with a mixture of siRNAs targeting highly conserved sequences of the 3B region and the 3D polymerase gene in all FMDV serotypes resulted in nearly 100% suppression of virus growth.

In another approach, siRNAs were designed to specifically target the viral VP1 gene, which plays a key role in virus attachment. This resulted in a nearly 90% reduction in FMDV VP1 expression and conferred resistance to FMDV challenge in cultured cells which are susceptible to this virus [92]. Encouragingly,

pretreatment with siRNAs before infection made suckling mice significantly less susceptible to FMDV, and expression of siRNAs directed against the viral nonstructural protein 2B clearly inhibited virus replication in infected porcine cells [93].

Another RNAi target of agricultural interest is the bovine viral diarrhea virus (BVDV), an ubiquitously occurring pathogen that affects cattle herds worldwide resulting in respiratory disorders and increased susceptibility to other pathogens [94]. Lambeth and his group demonstrated that BVDV replication in bovine cells can be efficiently suppressed by RNAi [95]. They transfected shRNA expression vectors and siRNAs targeting the 5′ nontranslated region (NTR) and the region encoding the C protein of the viral genome into MDBK cells. After challenging with BVDV, they detected reduced virus titres by both siRNA and shRNA-mediated RNAi.

Farm animals, in particular swine and poultry, serve as key links between the natural reservoir of influenza A viruses and epidemics and pandemics in human populations. Due to repeated reassortment or mixing of RNA segments between influenza viruses from different species, virulent strains emerge periodically and often lead to devastating human catastrophes [96]. However, the emergence of the RNAi technology has opened many new options for preventing influenza virus infections in animals.

In initial studies, a set of siRNAs specific for conserved regions of the influenza virus genome could potently inhibit virus production in MDCK cells and embryonated chicken eggs [97]. In subsequent approaches, this strategy was extended to an established animal model of influenza infections by two independent groups. Tompkins and colleagues used siRNAs for targeting highly conserved regions of the viral nucleoprotein (NP) and acidic polymerase (PA). After administration of influenza virus-specific siRNAs via hydrodynamic i.v. injection [98], BALB/c mice were infected intranasally with influenza A/H1N1. Virus titre in lung homogenates were significantly reduced in siRNAs-treated mice when compared to control mice 48 h p.i [99]. In addition, they demonstrated that influenza-specific siRNA treatment can protect mice from otherwise lethal virus challenges.

Ge and coworkers administered influenza virus-specific siRNAs intravenously along with lentiviral shRNA expression vectors into C57BL/6 mice. They demonstrated that siRNAs as well as shRNAs can reduce influenza virus production in the lung when given either before or after virus infection and that the simultaneous use of two or more siRNAs specific for different virus genes resulted in a more severe reduction of virus titres [100].

A promising approach for the generation of influenza-resistant livestock was published by Wise and colleagues [101]. They used shRNA expression vectors, targeting the viral NP and PA gene for lentiviral-mediated generation of transgenic mice. Expression of the siRNAs was confirmed by an RNAse protection assay, and thus far, stable transmission of the transgene was observed up to the third generation. Currently, transgenic mice are mated to generate homozygous lines for delivering the final proof for influenza virus resistance in vivo.

Recently the generation of transgenic chicken expressing a shRNA molecule able to inhibit influenza virus polymerase activity [115] was reported. Although the transgenic chicken did not exhibit a higher resistance to high challenge doses of H5N1, a highly pathogenic avian influenza virus, they showed strongly reduced transmission of the infection to transgenic and even non-transgenic birds housed in direct contact with them, demonstrating that this strategy may be used to prevent transmission and propagation of an infection at the flock level.

Future Directions

The past decade was dominated by large-scale and high throughput nucleic acid analyses allowing comparative genome sequencing and expression profiling projects. The comprehensive and ongoing analysis of the huge data sets led to the need for an updated definition of the term ‘gene’ and the introduction of the term ‘epigenetics.’ Taking into account that Mendel’s and Morgan’s elements of heredity include multifunctional protein coding, structural, regulatory, and RNAs of unknown functions and gene regulation is more complex than previously assumed, the ‘gene’ is suggested to be ‘a union of genomic sequences encoding a coherent set of potentially overlapping functional products’ [102] and ‘epigenetics’ is defined to describe ‘stably inheritable phenotypes resulting from changes in

a chromosome without alterations of the DNA-sequence' [103]. The future challenge of the postgenomic era is subsumed as integrative, quantitative, and/or systems biology. 'Systems biology is the comprehensive and quantitative analysis of the interactions between all of the components of biological systems over time' [104]. 'Systems biology involves an iterative cycle, in which emerging biological problems drive the development of new technologies and computational tools' [104]. The further understanding of disease mechanisms also depends on these emerging disciplines.

The ongoing genome sequencing programs for various animal species and the increasing densities of SNP arrays will lead to the discovery of new QTLs underlying economically important traits such as disease resistance and susceptibility. In addition, complete genome sequences of many disease-causing pathogens are becoming available. Hence, genome data on host intrinsic factors and host–pathogen interactions causing disease can be used to increase the health of individuals or populations. Conventional breeding and genomic selection will increasingly benefit from the natural variations identified among the populations. This can be supplemented with gene transfer technologies allowing a more targeted approach toward desired animal breeding without the limitation of species barriers.

The future of transgene technologies is dependent on the simplification of the gene delivery systems along with targeted manipulation of animal genomes. The former aim is achieved by using lentiviral vectors which are highly efficient for domesticated animals including poultry [105, 106] and pets [107]. Gene targeting in species other than mice is limited as embryonic stem (ES) cells of farm or pet animals are unavailable and gene targeting via homologous recombination of embryonic and somatic cells and subsequent nuclear transfer is highly inefficient. However, the advent of the RNAi-technology offers new possibilities for specific gene targeting in animal species and will have a huge impact on transgenesis in the near future. Furthermore, the zinc finger nuclease (ZNF) technology has shown to be an attractive alternative to ES cell targeting and nuclear transfer technology [108] and was already applied successfully for targeted gene disruption in rats [109].

For further reading concerning the use of ZFN-technology in farm animals we refer to Kues and Niemann [116]. In addition, site-directed mutagenesis of genomes can be achieved by TALENs (transcription activator-like effector nucleases) which were originally identified in plant pathogens and recently were successfully used to generate knockout rats [117]. These site-specific nucleases may complement/enlarge the well established ZFN-technology for efficient gene targeting in livestock [118].

Last but not least, the cross-species generation of pluripotent/embryonic cell lines has gained new impetus through the induced pluripotent stem cell (iPS) technology, i.e., the reprogramming of somatic cells making them capable of embryogenesis (reviewed in [110]) and the recent isolation of authentic embryonic stem cells from rat blastocysts by novel culture conditions [111, 112]. In the future, animal transgenetics and animal disease resistance will be important in basic research and in the understanding of disease mechanisms. Bridging the gap between model and man by generating transgenic animals is fundamental to the development of novel therapeutics and disease prevention strategies.

Increased availability of genomic information of livestock species along with more sophisticated transgenic tools offers the potential to generate animal models to combat livestock diseases to a larger extent than ever before. However, animal geneticists/scientists must consider several important aspects. (1) The dissemination of the trait of interest such as disease resistance, introduced by a transgene will neither be simple nor fast, therefore cost–benefit calculations will probably decide on implementation of transgenic animals. For example, transgenic BSE-resistant cattle [77] will probably never gain importance in agriculture where culling is considered to meet demands with respect to cost efficiency and biosafety. However, BSE-resistant cattle may be engineered for the production of pharmaceuticals and therefore will have an enormous impact on providing safer drugs. (2) There is general public opposition to the use of transgenic livestock. However, if animals were resistant to zoonotic diseases, therefore resulting in reduced frequency of pandemics and epidemics such as those caused by influenza virus, attitudes of human societies might change [113]. In this context, recently, a trypansome lytic factor (TLF)

from baboons that protected mice both from animal and human-infective *Trypanosoma* subspecies was identified and suggested to be transferred to livestock [114]. Animal trypanosomiasis is one of the major parasitic diseases of livestock flocks and livestock are the major reservoir for human-pathogenic trypanosomes. (3) Scientists and society should clearly keep in mind that pathogens readily change their antigenic determinants and create novel subtypes to escape the 'resistant' host's immune system. Attempts to introduce resistance traits into animal populations either by conventional breeding or transgenesis should be subjected to thorough cost/detriment–benefit analyses.

Bibliography

1. de Koning DJ, Archibald A, Haley CS (2007) Livestock genomics: bridging the gap between mice and men. Trends Biotechnol 25:483–489
2. Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10:381–391
3. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet JC, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat Genet 40:449–454
4. Hammer RE, Pursel VG, Rexroad CE Jr, Wall RJ, Bolt DJ, Ebert KM, Palmiter RD, Brinster RL (1985) Production of transgenic rabbits, sheep and pigs by microinjection. Nature 315:680–683
5. Brem G, Brenig B, Goodmann HM, Selden RC, Graf F, Kruff B, Springman K, Hondele J, Meyer J, Winnacker E-L, Kräußlich H (1985) Production of transgenic mice, rabbits and pigs by microinjection into pronuclei. Reprod Domest Anim 20:251–252
6. Pursel VG, Hammer RE, Bolt DJ, Palmiter RD, Brinster RL (1990) Integration, expression and germ-line transmission of growth-related genes in pigs. J Reprod Fertil Suppl 41:77–87
7. Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce de Leon FA, Robl JM (1998) Cloned transgenic calves produced from nonquiescent fetal fibroblasts. Science 280:1256–1258
8. Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH (1997) Viable offspring derived from fetal and adult mammalian cells. Nature 385:810–813
9. Campbell KH, McWhir J, Ritchie WA, Wilmut I (1996) Sheep cloned by nuclear transfer from a cultured cell line. Nature 380:64–66
10. Schnieke AE, Kind AJ, Ritchie WA, Mycock K, Scott AR, Ritchie M, Wilmut I, Colman A, Campbell KH (1997) Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. Science 278:2130–2133
11. Hofmann A, Kessler B, Ewerling S, Weppert M, Vogg B, Ludwig H, Stojkovic M, Boelhauve M, Brem G, Wolf E, Pfeifer A (2003) Efficient transgenesis in farm animals by lentiviral vectors. EMBO Rep 4:1054–1060
12. Whitelaw CB, Radcliffe PA, Ritchie WA, Carlisle A, Ellard FM, Pena RN, Rowe J, Clark AJ, King TJ, Mitrophanous KA (2004) Efficient generation of transgenic pigs using equine infectious anaemia virus (EIAV) derived vector. FEBS Lett 571:233–236
13. Golding MC, Long CR, Carmell MA, Hannon GJ, Westhusin ME (2006) Suppression of prion protein in livestock by RNA interference. Proc Natl Acad Sci USA 103:5285–5290
14. Muller M, Brem G (1994) Transgenic strategies to increase disease resistance in livestock. Reprod Fertil Dev 6:605–613
15. Muller M, Brem G (1996) Intracellular, genetic or congenital immunisation–transgenic approaches to increase disease resistance of farm animals. J Biotechnol 44:233–242
16. Muller M, Brem G (1998) Transgenic approaches to the increase of disease resistance in farm animals. Rev Sci Tech 17:365–378
17. Whitelaw CB, Sang HM (2005) Disease-resistant genetically modified animals. Rev Sci Tech 24:275–283
18. Haller O, Arnheiter H, Gresser I, Lindenmann J (1981) Virus-specific interferon action. Protection of newborn Mx carriers against lethal infection with influenza virus. J Exp Med 154:199–203
19. Staeheli P, Haller O, Boll W, Lindenmann J, Weissmann C (1986) Mx protein: constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. Cell 44:147–158
20. Arnheiter H, Skuntz S, Noteborn M, Chang S, Meier E (1990) Transgenic mice with intracellular immunity to influenza virus. Cell 62:51–61
21. Garber EA, Chute HT, Condra JH, Gotlib L, Colonno RJ, Smith RG (1991) Avian cells expressing the murine Mx1 protein are resistant to influenza virus infection. Virology 180:754–762
22. Muller M, Brenig B, Winnacker EL, Brem G (1992) Transgenic pigs carrying cDNA copies encoding the murine Mx1 protein which confers resistance to influenza virus infection. Gene 121:263–270
23. Ko JH, Jin HK, Asano A, Takada A, Ninomiya A, Kida H, Hokiyama H, Ohara M, Tsuzuki M, Nishibori M, Mizutani M, Watanabe T (2002) Polymorphisms and the differential antiviral activity of the chicken Mx gene. Genome Res 12:595–601
24. Thomas AV, Palm M, Broers AD, Zezafoun H, Desmecht DJ (2006) Genomic structure, promoter analysis, and expression of the porcine (Sus scrofa) Mx1 gene. Immunogenetics 58:383–389
25. Ellinwood NM, McCue JM, Gordy PW, Bowen RA (1998) Cloning and characterization of cDNAs for a bovine (Bos taurus) Mx protein. J Interferon Cytokine Res 18:745–755
26. Whitelaw CB, Lillico SG, King T (2008) Production of transgenic farm animals by viral vector-mediated gene transfer. Reprod Domest Anim 43(Suppl 2):355–358

27. Pfeifer A (2004) Lentiviral transgenesis. Transgenic Res 13:513–522
28. Park F (2007) Lentiviral vectors: are they the future of animal transgenesis? Physiol Genomics 31:159–173
29. Aigner B, Klymiuk N, Wolf E (2010) Transgenic pigs for xenotransplantation: selection of promoter sequences for reliable transgene expression. Curr Opin Organ Transplant 15(2):201–206
30. Kues WA, Schwinzer R, Wirth D, Verhoeyen E, Lemme E, Herrmann D, Barg-Kues B, Hauser H, Wonigeit K, Niemann H (2006) Epigenetic silencing and tissue independent expression of a novel tetracycline inducible system in double-transgenic pigs. FASEB J 20:1200–1202
31. Hancock RE, Diamond G (2000) The role of cationic antimicrobial peptides in innate host defences. Trends Microbiol 8:402–410
32. Selsted ME, Ouellette AJ (2005) Mammalian defensins in the antimicrobial immune response. Nat Immunol 6:551–557
33. Lee DG, Hahm KS, Shin SY (2004) Structure and fungicidal activity of a synthetic antimicrobial peptide, P18, and its truncated peptides. Biotechnol Lett 26:337–341
34. Powers JP, Hancock RE (2003) The relationship between peptide structure and antibacterial activity. Peptides 24:1681–1691
35. Epand RM, Vogel HJ (1999) Diversity of antimicrobial peptides and their mechanisms of action. Biochim Biophys Acta 1462:11–28
36. Reddy KV, Yedery RD, Aranha C (2004) Antimicrobial peptides: premises and promises. Int J Antimicrob Agents 24:536–547
37. Yarus S, Rosen JM, Cole AM, Diamond G (1996) Production of active bovine tracheal antimicrobial peptide in milk of transgenic mice. Proc Natl Acad Sci USA 93:14118–14121
38. Reed WA, Elzer PH, Enright FM, Jaynes JM, Morrey JD, White KL (1997) Interleukin 2 promoter/enhancer controlled expression of a synthetic cecropin-class lytic peptide in transgenic mice and subsequent resistance to Brucella abortus. Transgenic Res 6:337–347
39. Erskine RJ, Walker RD, Bolin CA, Bartlett PC, White DG (2002) Trends in antibacterial susceptibility of mastitis pathogens during a seven-year period. J Dairy Sci 85:1111–1118
40. Kumar JK (2008) Lysostaphin: an antistaphylococcal agent. Appl Microbiol Biotechnol 80:555–561
41. Kerr DE, Plaut K, Bramley AJ, Williamson CM, Lax AJ, Moore K, Wells KD, Wall RJ (2001) Lysostaphin expression in mammary glands confers protection against staphylococcal infection in transgenic mice. Nat Biotechnol 19:66–70
42. Wall RJ, Powell AM, Paape MJ, Kerr DE, Bannerman DD, Pursel VG, Wells KD, Talbot N, Hawk HW (2005) Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection. Nat Biotechnol 23:445–451
43. Maga EA, Anderson GB, Huang MC, Murray JD (1994) Expression of human lysozyme mRNA in the mammary gland of transgenic mice. Transgenic Res 3:36–42
44. Maga EA, Shoemaker CF, Rowe JD, Bondurant RH, Anderson GB, Murray JD (2006) Production and processing of milk from transgenic goats expressing human lysozyme in the mammary gland. J Dairy Sci 89:518–524
45. Schukken YH, Wilson DJ, Welcome F, Garrison-Tikofsky L, Gonzalez RN (2003) Monitoring udder health and milk quality using somatic cell counts. Vet Res 34:579–596
46. Maga EA, Cullor JS, Smith W, Anderson GB, Murray JD (2006) Human lysozyme expressed in the mammary gland of transgenic dairy goats can inhibit the growth of bacteria that cause mastitis and the cold-spoilage of milk. Foodborne Pathog Dis 3:384–392
47. Lonnerdal B (2003) Nutritional and physiologic significance of human milk proteins. Am J Clin Nutr 77:1537S–1543S
48. Brundige DR, Maga EA, Klasing KC, Murray JD (2008) Lysozyme transgenic goats' milk influences gastrointestinal morphology in young pigs. J Nutr 138:921–926
49. Zhang J, Li L, Cai Y, Xu X, Chen J, Wu Y, Yu H, Yu G, Liu S, Zhang A, Cheng G (2008) Expression of active recombinant human lactoferrin in the milk of transgenic goats. Protein Expr Purif 57:127–135
50. Salzman NH, Ghosh D, Huttner KM, Paterson Y, Bevins CL (2003) Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin. Nature 422:522–526
51. Panyutich A, Shi J, Boutz PL, Zhao C, Ganz T (1997) Porcine polymorphonuclear leukocytes generate extracellular microbicidal activity by elastase-mediated activation of secreted proprotegrins. Infect Immun 65:978–985
52. Cheung QC, Turner PV, Song C, Wu D, Cai HY, MacInnes JI, Li J (2008) Enhanced resistance to bacterial infection in protegrin-1 transgenic mice. Antimicrob Agents Chemother 52:1812–1819
53. Mauch C, Bilkei G (2004) Actinobacillus suis, a potential cause of abortion in gilts and low parity sows. Vet J 168:186–187
54. McManus MT, Sharp PA (2002) Gene silencing in mammals by small interfering RNAs. Nat Rev Genet 3:737–747
55. Gavora J (1996) Resistance to livestock to viruses: mechanisms and strategies for genetic engineering. Genet Sel Evol 28:385–414
56. Salter DW, Crittenden LB (1989) Artificial insertion of a dominant gene for resistance to avian leukosis virus into the germ line of the chicken. Theor Appl Genet 77:457–461
57. Clements JE, Wall RJ, Narayan O, Hauer D, Schoborg R, Sheffer D, Powell A, Carruth LM, Zink MC, Rexroad CE (1994) Development of transgenic sheep that express the visna virus envelope gene. Virology 200:370–380
58. Ono E, Amagai K, Taharaguchi S, Tomioka Y, Yoshino S, Watanabe Y, Cherel P, Houdebine LM, Adam M, Eloit M, Inobe M, Uede T (2004) Transgenic mice expressing a soluble form of porcine nectin-1/herpesvirus entry mediator C as a model for pseudorabies-resistant livestock. Proc Natl Acad Sci USA 101:16150–16155
59. Mulder WA, Pol JM, Gruys E, Jacobs L, De Jong MC, Peeters BP, Kimman TG (1997) Pseudorabies virus infections in pigs. Role of viral proteins in virulence, pathogenesis and transmission. Vet Res 28:1–17
60. Ono E, Tomioka Y, Watanabe Y, Amagai K, Taharaguchi S, Glenisson J, Cherel P (2006) The first immunoglobulin-like domain of porcine nectin-1 is sufficient to confer resistance

to pseudorabies virus infection in transgenic mice. Arch Virol 151:1827–1839
61. Tomioka Y, Morimatsu M, Amagai K, Kuramochi M, Watanabe Y, Kouda S, Wada T, Kuboki N, Ono E (2009) Fusion protein consisting of the first immunoglobulin-like domain of porcine nectin-1 and Fc portion of human IgG1 provides a marked resistance against pseudorabies virus infection to transgenic mice. Microbiol Immunol 53:8–15
62. Yoshida K, Tomioka Y, Kase S, Morimatsu M, Shinya K, Ohno S, Ono E (2008) Microphthalmia and lack of vitreous body in transgenic mice expressing the first immunoglobulin-like domain of nectin-1. Graefes Arch Clin Exp Ophthalmol 246:543–549
63. Lo D, Pursel V, Linton PJ, Sandgren E, Behringer R, Rexroad C, Palmiter RD, Brinster RL (1991) Expression of mouse IgA by transgenic mice, pigs and sheep. Eur J Immunol 21:1001–1006
64. Weidle UH, Lenz H, Brem G (1991) Genes encoding a mouse monoclonal antibody are expressed in transgenic mice, rabbits and pigs. Gene 98:185–191
65. Castilla J, Pintado B, Sola I, Sanchez-Morgado JM, Enjuanes L (1998) Engineering passive immunity in transgenic mice secreting virus-neutralizing antibodies in milk. Nat Biotechnol 16:349–354
66. Kolb AF, Pewe L, Webster J, Perlman S, Whitelaw CB, Siddell SG (2001) Virus-neutralizing monoclonal antibody expressed in milk of transgenic mice provides full protection against virus-induced encephalitis. J Virol 75:2803–2809
67. Saif LJ, Wheeler MB (1998) WAPping gastroenteritis with transgenic antibodies. Nat Biotechnol 16:334–335
68. Aguzzi A, Calella AM (2009) Prions: protein aggregation and infectious diseases. Physiol Rev 89:1105–1152
69. Aguzzi A, Sigurdson C, Heikenwaelder M (2008) Molecular mechanisms of prion pathogenesis. Annu Rev Pathol 3:11–40
70. Bueler H, Aguzzi A, Sailer A, Greiner RA, Autenried P, Aguet M, Weissmann C (1993) Mice devoid of PrP are resistant to scrapie. Cell 73:1339–1347
71. Bueler H, Fischer M, Lang Y, Bluethmann H, Lipp HP, DeArmond SJ, Prusiner SB, Aguet M, Weissmann C (1992) Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. Nature 356:577–582
72. Prusiner SB, Groth D, Serban A, Koehler R, Foster D, Torchia M, Burton D, Yang SL, DeArmond SJ (1993) Ablation of the prion protein (PrP) gene in mice prevents scrapie and facilitates production of anti-PrP antibodies. Proc Natl Acad Sci USA 90:10608–10612
73. Weissmann C, Bueler H, Fischer M, Sailer A, Aguzzi A, Aguet M (1994) PrP-deficient mice are resistant to scrapie. Ann NY Acad Sci 724:235–240
74. Denning C, Burl S, Ainslie A, Bracken J, Dinnyes A, Fletcher J, King T, Ritchie M, Ritchie WA, Rollo M, de Sousa P, Travers A, Wilmut I, Clark AJ (2001) Deletion of the alpha (1, 3) galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. Nat Biotechnol 19:559–562
75. Yu G, Chen J, Yu H, Liu S, Xu X, Sha H, Zhang X, Wu G, Xu S, Cheng G (2006) Functional disruption of the prion protein gene in cloned goats. J Gen Virol 87:1019–1027
76. Yu G, Chen J, Xu Y, Zhu C, Yu H, Liu S, Sha H, Xu X, Wu Y, Zhang A, Ma J, Cheng G (2009) Generation of goats lacking prion protein. Mol Reprod Dev 76:3
77. Richt JA, Kasinathan P, Hamir AN, Castilla J, Sathiyaseelan T, Vargas F, Sathiyaseelan J, Wu H, Matsushita H, Koster J, Kato S, Ishida I, Soto C, Robl JM, Kuroiwa Y (2007) Production of cattle lacking prion protein. Nat Biotechnol 25:132–138
78. Kuroiwa Y, Kasinathan P, Matsushita H, Sathiyaselan J, Sullivan EJ, Kakitani M, Tomizuka K, Ishida I, Robl JM (2004) Sequential targeting of the genes encoding immunoglobulin-mu and prion protein in cattle. Nat Genet 36:775–780
79. Moore RC, Lee IY, Silverman GL, Harrison PM, Strome R, Heinrich C, Karunaratne A, Pasternak SH, Chishti MA, Liang Y, Mastrangelo P, Wang K, Smit AF, Katamine S, Carlson GA, Cohen FE, Prusiner SB, Melton DW, Tremblay P, Hood LE, Westaway D (1999) Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. J Mol Biol 292:797–817
80. Rossi D, Cozzio A, Flechsig E, Klein MA, Rulicke T, Aguzzi A, Weissmann C (2001) Onset of ataxia and Purkinje cell loss in PrP null mice inversely correlated with Dpl level in brain. EMBO J 20:694–702
81. Tobler I, Gaus SE, Deboer T, Achermann P, Fischer M, Rulicke T, Moser M, Oesch B, McBride PA, Manson JC (1996) Altered circadian activity rhythms and sleep in mice devoid of prion protein. Nature 380:639–642
82. Collinge J, Whittington MA, Sidle KC, Smith CJ, Palmer MS, Clarke AR, Jefferys JG (1994) Prion protein is necessary for normal synaptic function. Nature 370:295–297
83. Melo EO, Canavessi AM, Franco MM, Rumpf R (2007) Animal transgenesis: state of the art and applications. J Appl Genet 48:47–61
84. Hannon GJ (2002) RNA interference. Nature 418:244–251
85. Daude N, Marella M, Chabry J (2003) Specific inhibition of pathological prion protein accumulation by small interfering RNAs. J Cell Sci 116:2775–2779
86. Pfeifer A, Eigenbrod S, Al-Khadra S, Hofmann A, Mitteregger G, Moser M, Bertsch U, Kretzschmar H (2006) Lentivector-mediated RNAi efficiently suppresses prion protein and prolongs survival of scrapie-infected mice. J Clin Invest 116:3204–3210
87. White MD, Farmer M, Mirabile I, Brandner S, Collinge J, Mallucci GR (2008) Single treatment with RNAi against prion protein rescues early neuronal dysfunction and prolongs survival in mice with prion disease. Proc Natl Acad Sci USA 105:10238–10243
88. Berkhout B, Haasnoot J (2006) The interplay between virus infection and the cellular RNA interference machinery. FEBS Lett 580:2896–2902
89. Colbere-Garapin F, Blondel B, Saulnier A, Pelletier I, Labadie K (2005) Silencing viruses by RNA interference. Microbes Infect 7:767–775
90. Grubman MJ, Baxt B (2004) Foot-and-mouth disease. Clin Microbiol Rev 17:465–493

91. Kahana R, Kuznetzova L, Rogel A, Shemesh M, Hai D, Yadin H, Stram Y (2004) Inhibition of foot-and-mouth disease virus replication by small interfering RNA. J Gen Virol 85:3213–3217
92. Chen W, Yan W, Du Q, Fei L, Liu M, Ni Z, Sheng Z, Zheng Z (2004) RNA interference targeting VP1 inhibits foot-and-mouth disease virus replication in BHK-21 cells and suckling mice. J Virol 78:6900–6907
93. de Los Santos T, Wu Q, de Avila Botton S, Grubman MJ (2005) Short hairpin RNA targeted to the highly conserved 2B nonstructural protein coding region inhibits replication of multiple serotypes of foot-and-mouth disease virus. Virology 335:222–231
94. Lindberg AL (2003) Bovine viral diarrhoea virus infections and its control. A review. Vet Q 25:1–16
95. Lambeth LS, Moore RJ, Muralitharan MS, Doran TJ (2007) Suppression of bovine viral diarrhea virus replication by small interfering RNA and short hairpin RNA-mediated RNA interference. Vet Microbiol 119:132–143
96. Cox NJ, Subbarao K (2000) Global epidemiology of influenza: past and present. Annu Rev Med 51:407–421
97. Ge Q, McManus MT, Nguyen T, Shen CH, Sharp PA, Eisen HN, Chen J (2003) RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. Proc Natl Acad Sci USA 100:2718–2723
98. Lewis DL, Hagstrom JE, Loomis AG, Wolff JA, Herweijer H (2002) Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. Nat Genet 32:107–108
99. Tompkins SM, Lo CY, Tumpey TM, Epstein SL (2004) Protection against lethal influenza virus challenge by RNA interference in vivo. Proc Natl Acad Sci USA 101:8682–8686
100. Ge Q, Filip L, Bai A, Nguyen T, Eisen HN, Chen J (2004) Inhibition of influenza virus production in virus-infected mice by RNA interference. Proc Natl Acad Sci USA 101:8676–8681
101. Wise TG, Schafer DS, Lowenthal JW, Doran TJ (2008) The use of RNAi and transgenics to develop viral disease resistant livestock. Dev Biol (Basel) 132:377–382
102. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, post-ENCODE? History and updated definition. Genome Res 17:669–681
103. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. Genes Dev 23:781–783
104. Zak DE, Aderem A (2009) Systems biology of innate immunity. Immunol Rev 227:264–282
105. McGrew MJ, Sherman A, Ellard FM, Lillico SG, Gilhooley HJ, Kingsman AJ, Mitrophanous KA, Sang H (2004) Efficient production of germline transgenic chickens using lentiviral vectors. EMBO Rep 5:728–733
106. Scott BB, Lois C (2005) Generation of tissue-specific transgenic birds with lentiviral vectors. Proc Natl Acad Sci USA 102:16443–16447
107. Hong SG, Kim MK, Jang G, Oh HJ, Park JE, Kang JT, Koo OJ, Kim T, Kwon MS, Koo BC, Ra JC, Kim DY, Ko C, Lee BC (2009) Generation of red fluorescent protein transgenic dogs. Genesis 47:314–322
108. Le Provost F, Lillico S, Passet B, Young R, Whitelaw B, Vilotte JL (2010) Zinc finger nuclease technology heralds a new era in mammalian transgenesis. Trends Biotechnol 28(3):134–141
109. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433
110. Belmonte JC, Ellis J, Hochedlinger K, Yamanaka S (2009) Induced pluripotent stem cells and reprogramming: seeing the science through the hype. Nat Rev Genet 10:878–883
111. Buehr M, Meek S, Blair K, Yang J, Ure J, Silva J, McLay R, Hall J, Ying QL, Smith A (2008) Capture of authentic embryonic stem cells from rat blastocysts. Cell 135:1287–1298
112. Li P, Tong C, Mehrian-Shai R, Jia L, Wu N, Yan Y, Maxson RE, Schulze EN, Song H, Hsieh CL, Pera MF, Ying QL (2008) Germline competent embryonic stem cells derived from rat blastocysts. Cell 135:1299–1310
113. Chen J, Chen SC, Stern P, Scott BB, Lois C (2008) Genetic strategy to prevent influenza virus infections in animals. J Infect Dis 197(Suppl 1):S25–S28
114. Thomson R, Molina-Portela P, Mott H, Carrington M, Raper J (2009) Hydrodynamic gene delivery of baboon trypanosome lytic factor eliminates both animal and human-infective African trypanosomes. Proc Natl Acad Sci USA 106:19509–19514
115. Lyall J, Irvine RM, Sherman A, McKinley TJ, Núnez A, Purdie A, Outtrim L, Brown IH, Rolleston-Smith G, Sang H, Tiley L (2011) Suppression of avian influenza transmission in genetically modified chickens. Science 331:223–226
116. Kues WA, Niemann H (2011) Advances in farm transgenesis. PREVET, doi:10.1016
117. Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ (2011) Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol 29:695–696
118. DeFrancesco L (2011) Move over ZFNs. Nat Biotechnol 29:681–684

Distribution Systems, Substations, and Integration of Distributed Generation

John D. McDonald[1], Bartosz Wojszczyk[2]
Byron Flynn[2], Ilia Voloh[2]
[1]GE Energy, Digital Energy, Atlanta, GA, USA
[2]GE Energy, Digital Energy, Norcross, GA, USA

Article Outline

Glossary
Definition of the Subject

Introduction
Distribution Systems
Substations
High Penetration of Distributed Generation and Its Impact on System Design and Operations
Future Directions
Bibliography

Glossary

Demand response Allows the management of customer consumption of electricity in response to supply conditions.

Distributed generation Electric energy that is distributed to the grid from many decentralized locations, such as from wind farms and solar panel installations.

Distribution grid The part of the grid dedicated to delivering electric energy directly to residential, commercial, and industrial electricity customers.

Distribution management system A smart grid automation technology that provides real time about the distribution network and allows utilities to remotely control devices in the grid.

Distribution substation Delivers electric energy to the distribution grid.

Distribution system The link from the distribution substation to the customer.

Renewable energy Energy from natural resources such as sunlight, wind, rain, tides, biofuels, and geothermal heat, which are naturally replenished.

Smart grid A modernization of the electricity delivery system so it monitors, protects, and automatically optimizes the operation of its interconnected elements.

Definition of the Subject

This entry describes the major components of the electricity distribution system – the distribution network, substations, and associated electrical equipment and controls – and how incorporating automated distribution management systems, devices, and controls into the system can create a "smart grid" capable of handling the integration of large amounts of distributed (decentralized) generation of sustainable, renewable energy sources.

Introduction

Distributed generation (DG) or decentralized generation is not a new industry concept. In 1882, Thomas Edison built his first commercial electric plant – "Pearl Street." The Pearl Street station provided 110 V direct current (DC) electric power to 59 customers in lower Manhattan. By 1887, there were 121 Edison power stations in the United States delivering DC electricity to customers. These early power plants ran on coal or water. Centralized power generation became possible when it was recognized that alternating current (AC) electricity could be transported at relatively low costs with reduced power losses across great distances by taking advantage of the ability to raise the voltage at the generation station and lower the voltage near customer loads. In addition, the concepts of improved system performance (system stability) and more effective generation asset utilization provided a platform for wide-area grid integration. Recently, there has been a rapidly growing interest in wide deployment of distributed generation, which is electricity distributed to the grid from a variety of decentralized locations. Commercially available technologies for distributed generation are based on wind turbines, combustion engines, micro- and mini-gas turbines, fuel cells, photovoltaic (solar) installations, low-head hydro units, and geothermal systems.

Deregulation of the electric utility industry, environmental concerns associated with traditional fossil fuel generation power plants, volatility of electric energy costs, federal and state regulatory support of "green" energy, and rapid technological developments all support the proliferation of distributed generation in electric utility systems. The growing rate of DG deployment also suggests that alternative energy-based solutions will play an increasingly important role in the smart grid and modern utility.

Large-scale implementation of distributed generation can lead to the evolution of the distribution network from a "passive" (local/limited automation, monitoring, and control) system to an "active" (global/integrated, self-monitoring, semiautomated) system that automatically responds to the various dynamics of the electric grid, resulting in higher efficiency, better load management, and fewer outages. However, distributed generation also poses a challenge for the design, operation, and management of the power grid because the network no longer

behaves as it once did. Consequently, the planning and operation of new systems must be approached differently, with a greater amount of attention paid to the challenges of an automated global system.

This entry describes the major components and interconnected workings of the electricity distribution system, and addresses the impact of large-scale deployment of distributed generation on grid design, reliability, performance, and operation. It also describes the distributed generation technology landscape, associated engineering and design challenges, and a vision of the modern utility.

Distribution Systems

Distribution systems serve as the link from the distribution substation to the customer. This system provides the safe and reliable transfer of electric energy to various customers throughout the service territory. Typical distribution systems begin as the medium-voltage three-phase circuit, typically about 30–60 kV, and terminate at a lower secondary three- or single-phase voltage typically below 1 kV at the customer's premise, usually at the meter.

Distribution feeder circuits usually consist of overhead and underground circuits in a mix of branching laterals from the station to the various customers. The circuit is designed around various requirements such as required peak load, voltage, distance to customers, and other local conditions such as terrain, visual regulations, or customer requirements. These various branching laterals can be operated in a radial configuration or as a looped configuration, where two or more parts of the feeder are connected together usually through a normally open distribution switch. High-density urban areas are often connected in a complex distribution underground network providing a highly redundant and reliable means connecting to customers. Most three-phase systems are for larger loads such as commercial or industrial customers. The three-phase systems are often drawn as one line as shown in the following distribution circuit drawing (Fig. 1) of three different types of circuits.

The secondary voltage in North America and parts of Latin America consists of a split single-phase service that provides the customer with 240 and 120 V, which the customer then connects to devices depending on their ratings. This is served from a three-phase distribution feeder normally connected in a Y configuration consisting of a neutral center conductor and a conductor for each phase, typically assigned a letter A, B, or C.

Single-phase customers are then connected by a small neighborhood distribution transformer connected from one of the phases to neutral, reducing the voltage from the primary feeder voltage to the secondary split service voltage. In North America, normally 10 or fewer customers are connected to a distribution transformer.

In most other parts of the world, the single-phase voltage of 220 or 230 V is provided directly from a larger neighborhood distribution transformer. This provides a secondary voltage circuit often serving hundreds of customers.

Figure 1 shows various substations and several feeders serving customers from those substations. In Fig. 1, the primary transformers are shown as blue boxes in the substation, various switches, breakers, or reclosers are shown as red (closed) or green (open) shapes, and fuses are shown as yellow boxes.

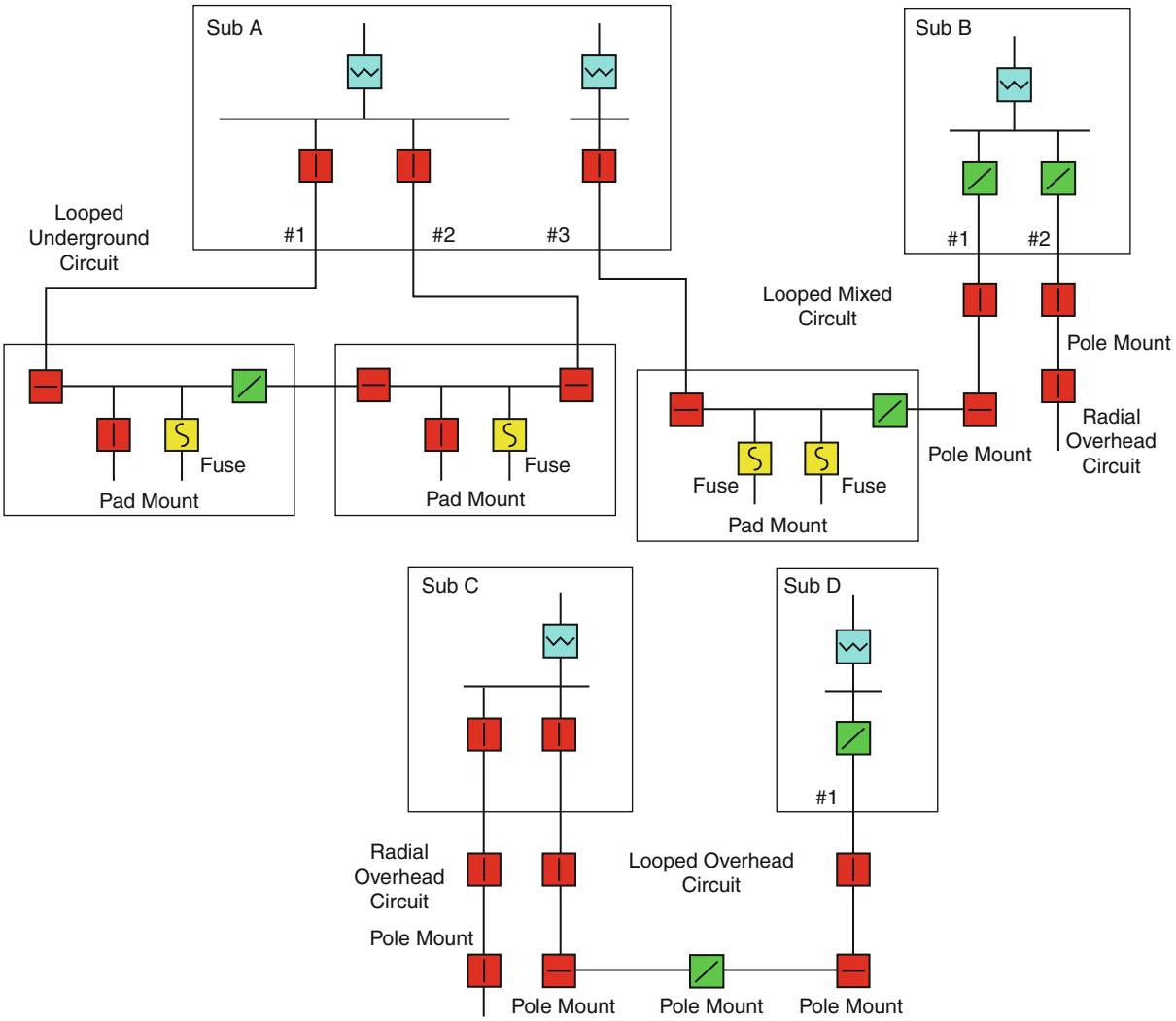
Distribution Devices

There are several distribution devices used to improve the safety, reliability, and power quality of the system. This section will review a few of those types of devices.

Switches: Distribution switches (Fig. 2) are used to disconnect various parts of the system from the feeder. These switches are manually, remotely, or automatically operated. Typically, switches are designed to break load current but not fault current and are used in underground circuits or tie switches.

Breakers: Like switches, distribution breakers are used to disconnect portions of the feeder. However, breakers have the ability to interrupt fault current. Typically, these are tied to a protective relay, which detects the fault conditions and issues the open command to the breaker.

Reclosers: These are a special type of breaker (Fig. 3), typically deployed only on overhead and are designed to reduce the outage times caused by momentary faults. These types of faults are caused by vegetation or temporary short circuits. During the reclose operation, the relay detects the fault, opens the switch, waits a few seconds, and issues a close. Many



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 1
Simple distribution system single line drawings

overhead distribution faults are successfully cleared and service is restored with this technique, significantly reducing outage times.

Capacitors: These are three-phase capacitors designed to inject volt amp reactives (VARs) into the distribution circuit, typically to help improve power factor or support system voltage (Fig. 4). They are operated in parallel with the feeder circuit and are controlled by a capacitor controller. These controllers are often connected to remote communications allowing for automatic or coordinated operation.

Fuses: These are standard devices used to protect portions of the circuit when a breaker is too expensive or too large. Fuses can be used to protect single-phase laterals off the feeder or to protect three-phase underground circuits.

Lightning arresters: These devices are designed to reduce the surge on the line when lightning strikes the circuit.

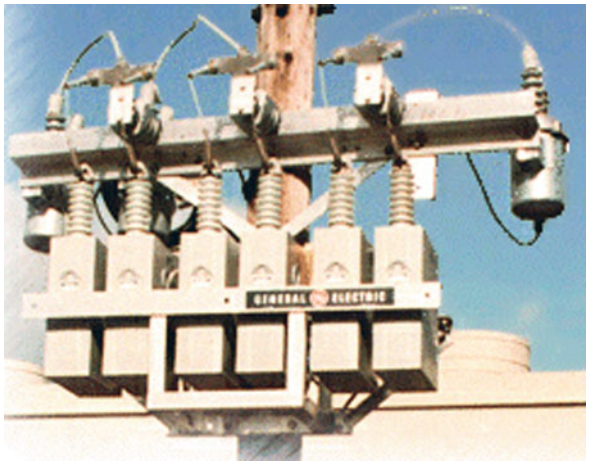
Automation Scheme: FDIR

The following description highlights an actual utility's FDIR automation scheme, their device decisions,

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 2
Distribution pad-mount switch

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 3
Distribution pole-mounted reclosing relay

functionality and system performance. Automation sequences include fault detection, localization, isolation, and load restoration (FDIR). These sequences will detect a fault, localize it to a segment of feeder, open the switches around the fault, and restore unfaulted sources via the substation and alternative sources as available. These algorithms work to safely minimize the fault duration and extent, significantly improving the SAIDI (system average interruption duration index) and SAIFI (system average



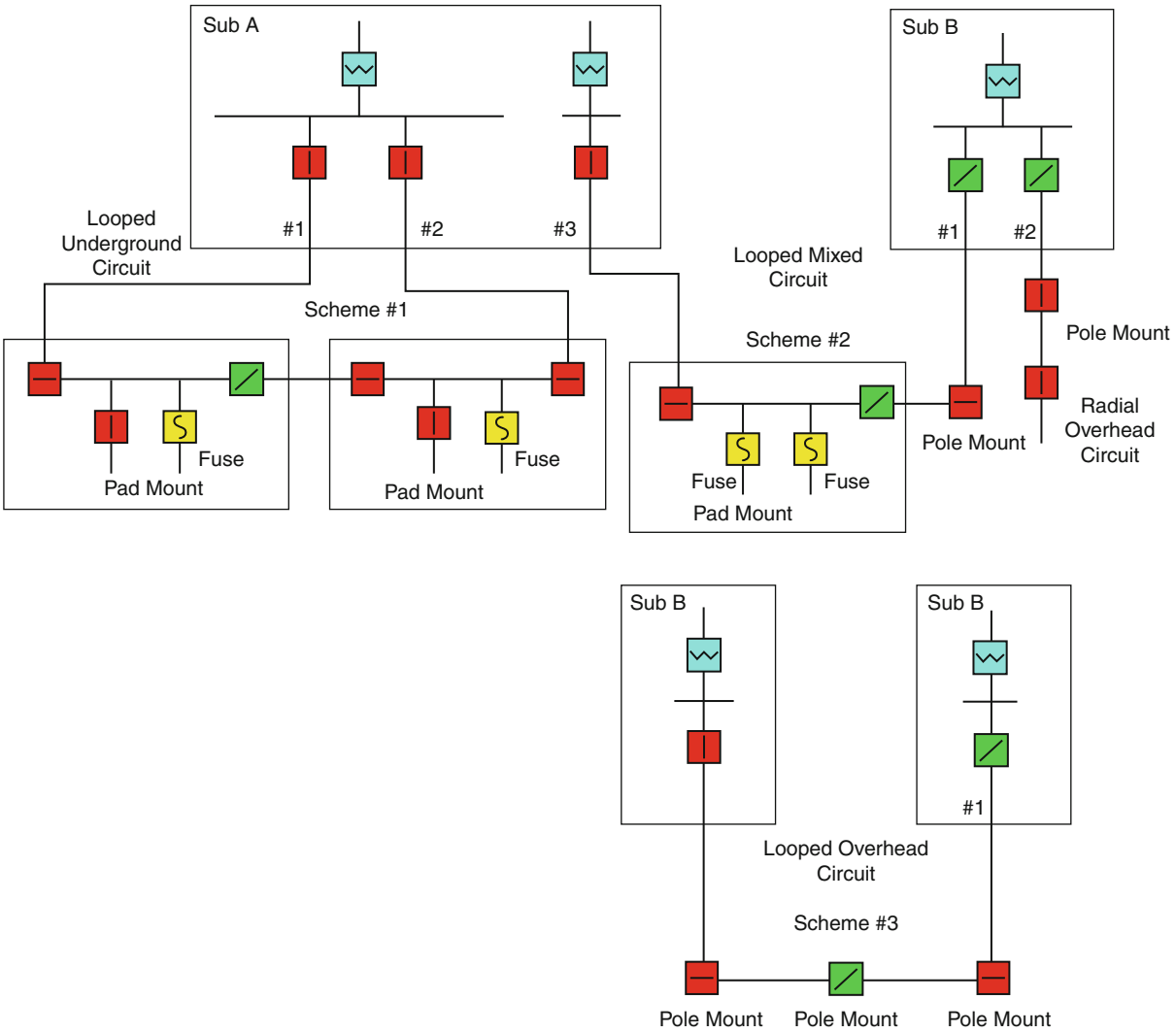
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 4
Distribution overhead 600 kVA capacitor

interruption frequency Index) performance metric for the customers on those feeders. An additional important sequence is the automatic check of equipment loading and thermal limits to determine whether load transfers can safely take place.

Modern systems communicate using a secure broadband Ethernet radio system, which provides significant improvement over a serial system, including supporting peer-to-peer communications, multiple access to tie switches, and remote access by communications and automation maintenance personnel. The communication system utilizes an internet protocol (IP)–based communication system with included security routines designed to meet the latest NERC (North American Electric Reliability Corporation) or the distribution grid operator's requirements.

Feeder circuits to be automated are typically selected because they have relatively high SAIDI indices serving high-profile commercial sites. Scheme 1 utilized two pad-mount switches connected to one substation. Scheme 2 consisted of a mix of overhead and underground with vault switchgear and a pole-mounted recloser. Scheme 3 was installed on overhead circuits with three pole-mounted reclosers.

Automation Schemes 1, 2 and 3 (Fig. 5) were designed to sense distribution faults, isolate the appropriate faulted line sections, and restore unfaulted circuit sections as available alternate source capacity permitted.



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 5
Distribution automation (DA) system single lines

Safety Safety is a critical piece of system operation. Each algorithm has several safety checks before any operation occurs. Before the scheme logic is initialized, a series of checks occur, including:

- Auto Restoration is enabled on a specific scheme – dispatchers do this via the distribution management system or SCADA system.
- Auto Restoration has not been disabled by a crew in the field via enable/disable switches at each device location.
- Auto Restoration has been reset – each scheme must be reset by the dispatcher after DA has operated and system has been restored to normal configuration.
- Communications Status – verifies that all necessary devices are on-line and communicating.
- Switch Position – verifies that each appropriate line switch is in the appropriate position (see Fig. 1).
- Voltages – checks that the appropriate buses/feeders are energized.
- Feeder Breaker Position – verifies the faulted feeder breaker has locked open and was opened only by

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 6
Pad-mount controller and pole-mount reclosing relay with enable/disable switches

a relay, not by SCADA or by the breaker control handle.

Prior to closing the tie switch and restoring customers in un-faulted sections, the following safety checks occur:

- Determine Pre-Fault Load – determine pre-fault load on un-faulted section of line.
- Compare Pre-Fault Load to Capacity – determine if alternate source can handle the un-faulted line section load.

After any DA algorithm executes:

- Notifies Dispatch of Status of DA System – success or failure of restoring load in un-faulted line sections.
- Reset is Necessary – algorithm is disabled until reset by dispatcher once the fault is repaired and the system is put back to normal configuration (see Fig. 1).

In summary, automation can occur *only* if these five conditions are true for every device on a scheme:

- Enable/Disable Switch is in enable position
- Local/Remote switch is in remote
- Breaker "hot-line" tag is off.
- Breaker opens from a relay trip and stays open for several seconds (that is, goes to lockout).
- Dispatch has reset the scheme(s) after the last automation activity.

Each pad-mounted or pole-mounted switch has a local enable/disable switch as shown in Fig. 6. Journeymen are to use these switches as the primary means of disabling a DA scheme before starting work on any of the six automated circuits or circuit breakers or any of the seven automated line switches.

FDIR System Components The automation system consists of controllers located in pad-mount switches, pole-mounted recloser controls (Fig. 7), and in substations (Fig. 8).

Pad-Mounted Controller The pad-mounted controller was selected according to the following criteria:

- Similar to existing substation controllers – simplifying configuration and overall compatibility
- Compatible with existing communications architecture
- Uses IEC 61131-3 programming
- Fault detection on multiple circuits
- Ethernet connection
- Supports multiple master stations
- Installed cost

The pad-mounted controller (Fig. 7) selected was an Ethernet-based controller that supported the necessary above requirements. The IEC 61131-3 programming languages include:

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 7
Typical pad-mount and pole-mount switches

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 8
Typical substation controller and vault switch

- Sequential Function Chart – describes operations of a sequential process using a simple graphical representation for the different steps of the process, and conditions that enable the change of active steps. It is considered to be the core of the controller configuration language, with other languages used to define the steps within the flowchart.
- Flowchart – a decision diagram composed of actions and tests programmed in structured text, instruction list, or ladder diagram languages. This is a proposed IEC 61131-3 language.
- Function Block Diagram – a graphic representation of many different types of equations. Rectangular boxes represent operators, with inputs on the left side of the box and outputs on the right. Custom function blocks may be created as well. Ladder diagram expressions may be a part of a function block program.
- Ladder Diagram – commonly referred to as "quick LD," the LD language provides a graphic representation of Boolean expressions. Function blocks can be inserted into LD programs.
- Structured Text – high-level structured language designed for expressing complex automation processes which cannot be easily expressed with graphic languages. Contains many expressions common to software programming languages (CASE, IF-THEN-ELSE, etc.). It is the default language for describing sequential function chart steps.
- Instruction List – a low-level instruction language analogous to assembly code.

Pole-Mounted Controller with Recloser The recloser controller was selected according to the following criteria:

- Similar requirements to pad-mounted controllers
- Control must provide needed analog and status outputs to DA remote terminal units (RTU)

Substation Controller The substation controller (Fig. 8) was selected per the following criteria:

- Similar to field controllers – simplifying configuration and overall compatibility
- Compatible with communications architecture
- Uses IEC 61131-3
- Ethernet and serial connections
- Supports multiple master stations including master station protocols
- Remote configuration is supported

Communications System

General The primary requirement of the communications system was to provide a secure channel between the various switches and the substation. The communication channel also needed to allow remote connection to the switchgear intelligent electronic devices (IEDs) for engineers and maintenance personnel. Additionally, the DA system also required the support for multiple substation devices to poll the controller at the tie switch. These requirements indicated the need for multi-channel or broadband radio.

Radio Communication Selection Criteria Primary considerations for selecting radio communications include:

- Security
- Supports remote configuration
- Broadband or multiple channels
- Compatible with multiple protocols
- From a major supplier
- Installed cost

The radio selected is a broadband radio operating over 900 MHz spread spectrum and 512 kbps of bandwidth. The wide-area network (WAN), Ethernet-based radio, supports the necessary protocols and provides multiple communications channels.

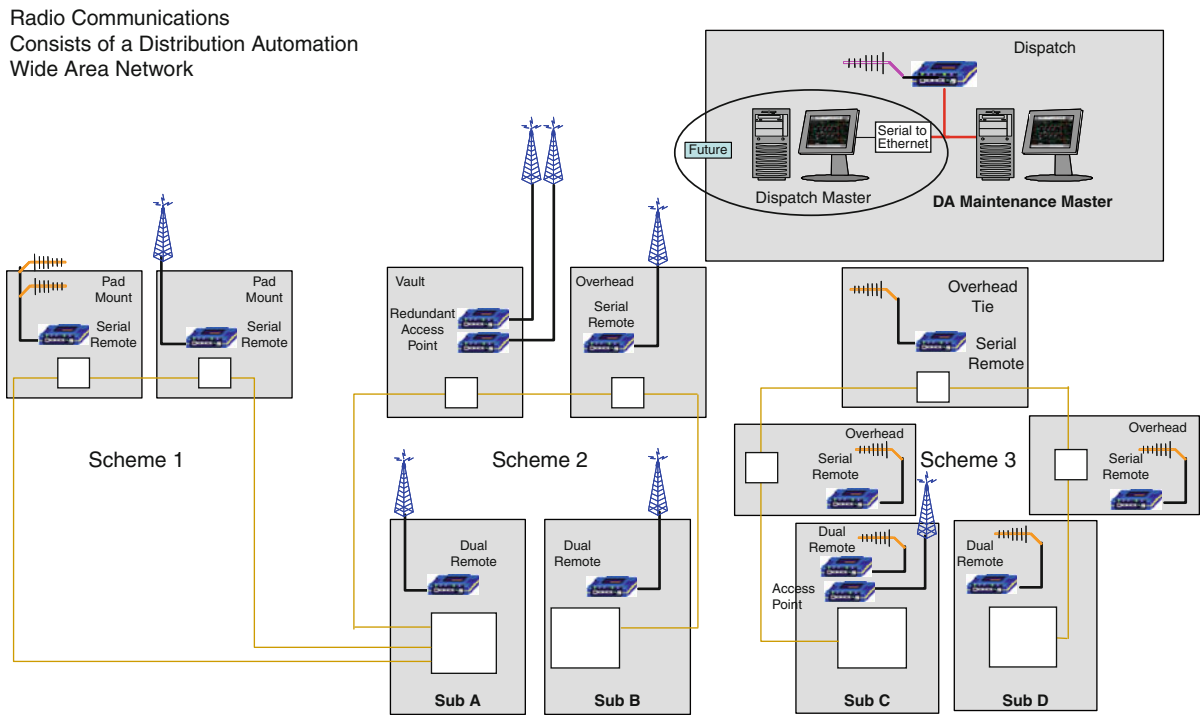
The communications network operates as a WAN providing the capability to communicate between any two points simply by plugging into the 10baseT communications port. A DA maintenance master was installed to communicate with the various controllers and to provide a detailed view of the DA system from the dispatch center. The DA system was also connected to the dispatch master station, which gives the dispatcher the ability to monitor and control the various DA algorithms and, in the future, typical SCADA control of the switches using distribution network protocol (DNP). (Since the dispatch master station currently does not support DNP over IP, a serial to Ethernet converter will be installed at the dispatch center to handle the conversion.) Figure 9 illustrates the communications architecture.

The radios communicate using point to multipoint with an access point radio operating as the base station radio and two types of remote radios, with serial or Ethernet. Some of the remote controllers only used DNP serial channels, requiring the radios to convert the serial connection to Ethernet. The remote Ethernet-based devices connect to the radios using a standard 10BaseT connection. Refer to Fig. 9.

For sites that require multiple DNP masters to connect to the serial controllers, the radios and the controllers have two serial connections. A new feature of the serial controllers is the support of point-to-point protocol (PPP). PPP provides a method for transmitting datagrams over serial point-to-point links. PPP contains three main components:

- A method for encapsulating datagrams over serial links. PPP uses the high-level data link control (HDLC) protocol as a basis for encapsulating datagrams over point-to-point links.
- An extensible Link Control Protocol (LCP) which is used to establish, configure, maintain, terminate, and test the data link connection.
- A family of network control protocols (NCP) for establishing and configuring different network layer protocols after LCP has established a connection. PPP is designed to allow the simultaneous use of multiple network layer protocols.

Establishing a PPP connection provides support for DNP multiple masters and a remote connection for maintenance over one serial communication line,



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 9
DA communications infrastructure

effectively providing full Ethernet functionality over a single serial channel.

Security The wireless system contains several security features. Table 1 outlines the threat and the security measures implemented in the radio to meet these threats.

Automation Functionality

Distribution Automation Schemes Distribution automation (DA) scheme operation is discussed in this section. All three schemes are configured the same, differing only in the type of midpoint and tie point switches used and whether the two sources are in the same or different substations. All three are set up as two-zone circuit pairs, with one tie point and two midpoints. Only one scheme will be shown (Fig. 10), as the others are analogous. Scheme 1 operates on the system shown in Fig. 10. It consists of two pad-mount switches and one substation. The pad-mount controllers communicate with the controller in the substation and the substation controller communicates with the DA maintenance master station and, in the future, the dispatch master.

Zone 1 Permanent Fault Before the algorithm operates, the safety checks occur as previously described. Refer to Figs. 11 and 12. If a permanent Zone 1 fault occurs (between switch 1 and substation CB14) and the algorithm is enabled and the logic has been initialized, the following actions occur:

1. After relaying locks out the substation breaker, the algorithm communicates with the field devices and the station protection relays to localize the fault.
2. Algorithm determines fault is between the substation and SW1.
3. Algorithm opens the circuit at SW1 connected to the incoming line from the substation, isolating the fault.
4. Algorithm gathers pre-fault load of section downstream of SW1 from the field devices.
5. Algorithm determines if capacity exists on alternate source and alternate feeder.

Distribution Systems, Substations, and Integration of Distributed Generation. Table 1 Security risk management

Security threat	900 MHz radio security
Unauthorized access to the backbone network through a foreign remote radio	Approved remotes list. Only those remotes included in the AP list will connect
"Rogue" AP, where a foreign AP takes control of some or all remote radios and thus remote devices	Approved AP List. A remote will only associate to those AP included in its local
Dictionary attacks, where a hacker runs a program that sequentially tries to break a password	Failed-login lockdown. After three tries, the transceiver ignores login requests for 5 min. Critical event reports (traps) are generated as well
Denial of service, where Remote radios could be reconfigured with bad parameters bringing the network down	• Remote login
	• Local console login
	• Disabled HTTP and Telnet to allow only local management services
Airwave searching and other hackers in parking lots, etc.	• 900 MHz FHSS does not talk over the air with standard 802.11b cards.
	• The transceiver cannot be put in a promiscuous mode.
	• Proprietary data framing
Eavesdropping, intercepting messages	128-bit encryption
Key cracking	Automatic rotating key algorithm
Replaying messages	128-bit encryption with rotating keys
Unprotected access to configuration via SNMPv1	Implement SNMPv3 secure operation
Intrusion detection	Provides early warning via SNMP through critical event reports (unauthorized, logging attempts, etc.)

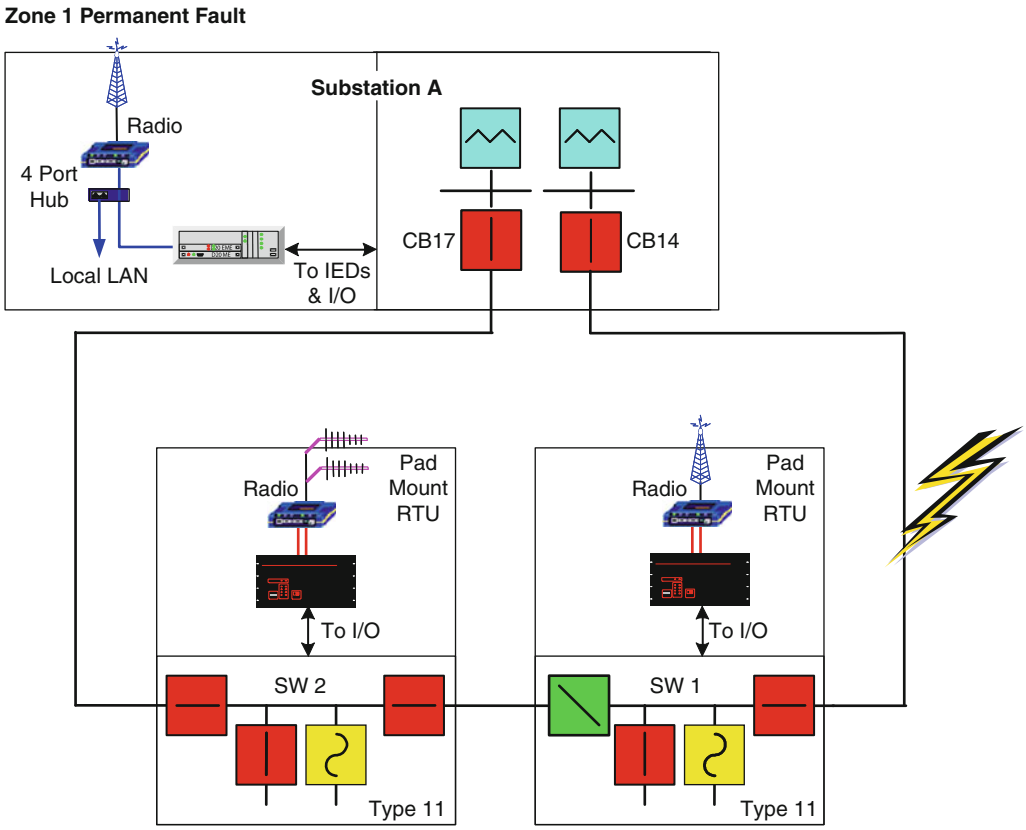
6. If so, algorithm closes the tie switch and backfeeds load, restoring customers on un-faulted line.
7. Reports successful operation to dispatch. The system is now as shown in Figs. 11 and 12, resulting in a reduction of SAIFI and SAIDI.

Zone 1 Permanent Fault: Load Too High to Safely Transfer In this case, a Zone 1 permanent fault occurs as shown in Fig. 13 and the previous example, except that this time loads are too high for the alternate source to accept load from the faulted feeder. Note the dispatch DA screens are descriptive and present information in plain language. Refer to Figs. 13 and 14.

Zone 2 Permanent Fault Depending on the type of the SW1 device (Fig. 15), the following actions occur:

If SW1 is a recloser (as in Schemes 2 and 3):

1. SW1 locks out in three shots. If SW1 is a pad-mount switch with no protection package (as in Scheme 1), the substation breaker goes to lockout. Fifty percent of CB11 customers remain in power.
2. This action occurs whether DA is enabled or disabled. That is, *existing circuit protection is unaffected by any DA scheme or logic.*
3. Safety checks are performed to ensure DA can safely proceed.
4. DA logic sees loss of voltage only beyond SW1 (recloser at lockout) and saw fault current through CB11 and SW1, so it recognizes that the line beyond SW1 is permanently faulted.
5. DA will not close into a faulted line, so the alternate source tie point (open point of SW2) remains open.
6. Customers between SW1 and SW2 lose power (about 50% of CB11 customers).



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 10
Scheme 1 architecture

If SW1 is switchgear (as in Scheme 1):

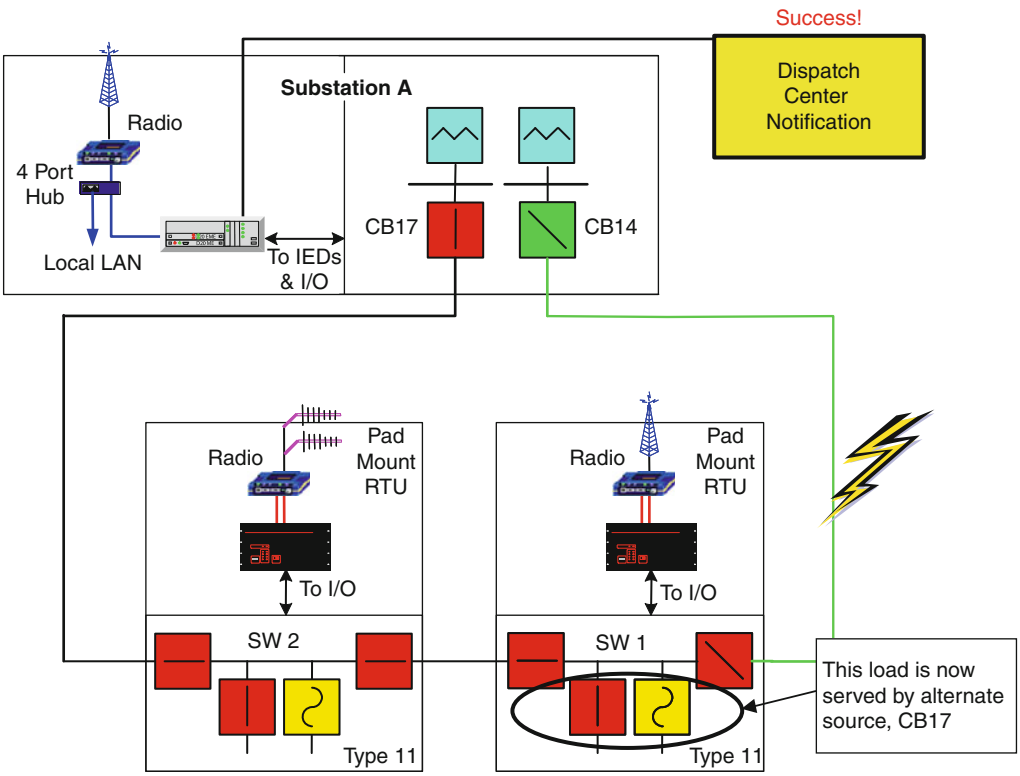
1. Substation circuit breaker, CB11, locks out in three shots.
2. This action occurs whether DA is enabled or disabled. That is, *existing circuit protection is unaffected by any DA scheme or logic.*
3. Safety checks are performed to ensure DA can safely proceed.
4. DA logic sees loss of voltage beyond CB11 (CB at lockout) and saw fault current through CB11 and SW1, so it recognizes that the line beyond (not before) SW1 is permanently faulted.
5. Fault is isolated by DA logic, sending open command to SW1.
6. DA logic recognizes line upstream of SW1 is good (fault current sensed at two devices), and closes CB11, heating up line to source side of open SW1. Power is now restored to 50% of customers.
7. DA will not close into a faulted line, so the alternate source tie point (open point of SW2) remains open (Figs. 16 and 17).

System Operation In 5 months of operation thus far the DA system has operated for 21 faults; all were Zone 2 faults on Scheme 3 (all downstream of the midpoint SW1). Three of those faults were permanent and took the line recloser SW1 to lockout. As a result, in 5 months, the DA pilot has saved 550 customers 6 h of power outage time (i.e., saved 3,300 customer hours lost) and eliminated 18 momentaries for those same 550 customers.

There have been no Zone 1 faults on any scheme; therefore, no load transfers to alternate sources have taken place.

Automation Scheme: Volt/VAR Control (VVC)

General The various loads along distribution feeders result in resistive (I^2R) and reactive (I^2X) losses in the



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 11
Scheme 1 architecture after successful DA operation

distribution system. If these losses are left uncompensated, an additional problem of declining voltage profile along the feeder will result. The most common solution to these voltage problems is to deploy voltage regulators at the station or along the feeder and/or a transformer LTC (load tap changers) on the primary station transformer; additional capacitors at the station and at various points on the feeder also provide voltage support and compensate the reactive loads. Refer to Fig. 18.

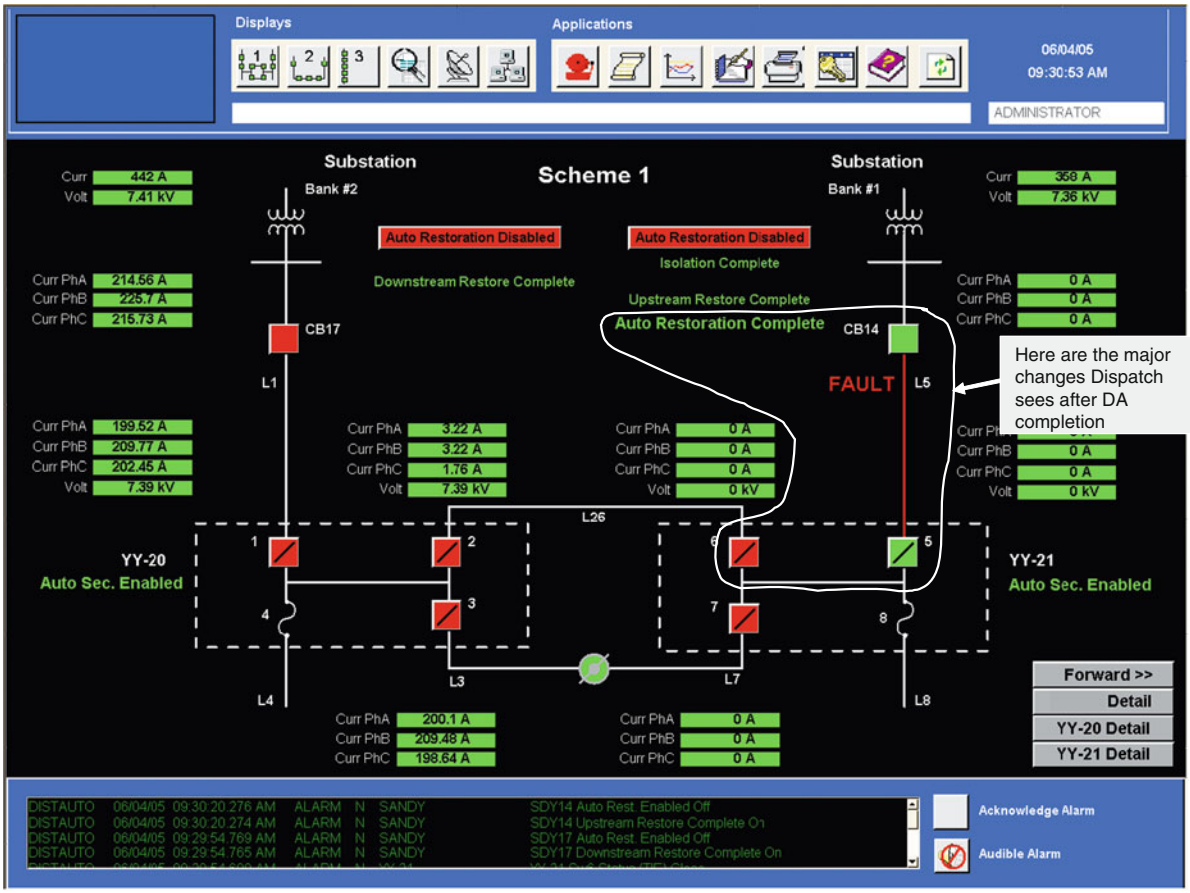
Many utilities are looking for additional benefits through improved voltage management. Voltage management can provide significant benefits through improved load management and improved voltage profile management.

The station Volt/VAR equipment consists of a primary transformer with either an LTC (Fig. 18) or a station voltage regulator and possibly station capacitors. The distribution feeders include line capacitors and possibly line voltage regulators.

The LTC is controlled by an automatic tap changer controller (ATC). The substation capacitors are controlled by a station capacitor controller (SCC), the distribution capacitors are controlled by an automatic capacitor controller (ACC), and the regulators are controlled by an automatic regulator controller (ARC). These controllers are designed to operate when local monitoring indicates a need for an operation including voltage and current sensing. Distribution capacitors are typically controlled by local power factor, load current, voltage, VAR flow, temperature, or the time (hour and day of week).

Some utilities have realized additional system benefits by adding communications to the substation, and many modern controllers support standard station communications protocols such as DNP.

This system (shown in Figs. 19 and 20) includes the ability to remotely monitor and manually control the volt/VAR resources, as well as the ability to provide integrated volt/VAR control (IVVC).



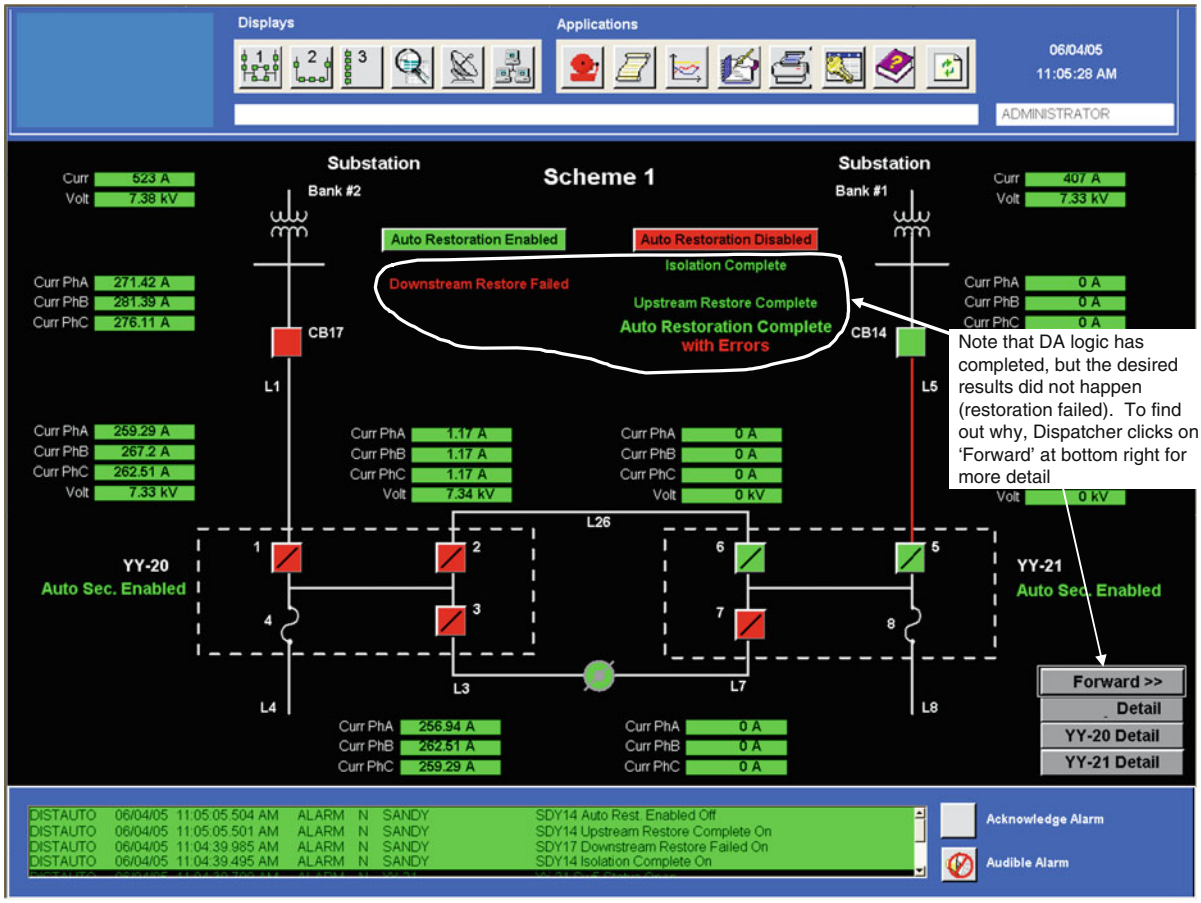
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 12
Dispatch notification of scheme 1 isolation/restoration success

Benefits of Volt/VAR Control (VVC) The VAR control systems can benefit from improved power factor and the ability to detect a blown fuse on the distribution capacitor. Studies and actual field data have indicated that systems often add an average of about 1 MVAR to each feeder. This can result in about a 2% reduction in the losses on the feeder.

Based on the assumptions, the benefits for line loss optimization that some utilities have calculated represent a significant cost-benefit payback. However, one of the challenges utilities face is that the cost and benefits are often disconnected. The utility's distribution business usually bears the costs for an IVVC system. The loss reduction benefits often initially flow to the transmission business and eventually to the ratepayer, since losses are covered in rates. Successful implementation of a loss reduction system will depend on helping align the costs with the benefits. Many utilities have successfully reconnected these costs and benefits of a Volt/VAR system through the rate process.

The voltage control systems can provide benefit from reduced cost of generation during peak times and improved capacity availability. This allows rate recovery to replace the loss of revenue created from voltage reduction when it is applied at times other than for capacity or economic reasons. The benefits for these programs will be highly dependent on the rate design, but could result in significant benefits.

There is an additional benefit from voltage reduction to the end consumer during off-peak times. Some utilities are approaching voltage reduction as a method to reduce load similar to a demand response (DR) program. Rate programs supporting DR applications are usually designed to allow the utility to



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 13
Dispatch notification of scheme 1 restoration failure

recapture lost revenue resulting from a decreased load. In simplified terms, the consumer would pay the utility equal to the difference between their normal rate and the wholesale price of energy based on the amount of load reduction. Figure 21 highlights the impact of voltage as a load management tool.

This chart contains real data from a working feeder utilizing Volt/VAR control. As the chart indicates, with VVC, the feeder voltage profile is flatter and lower.

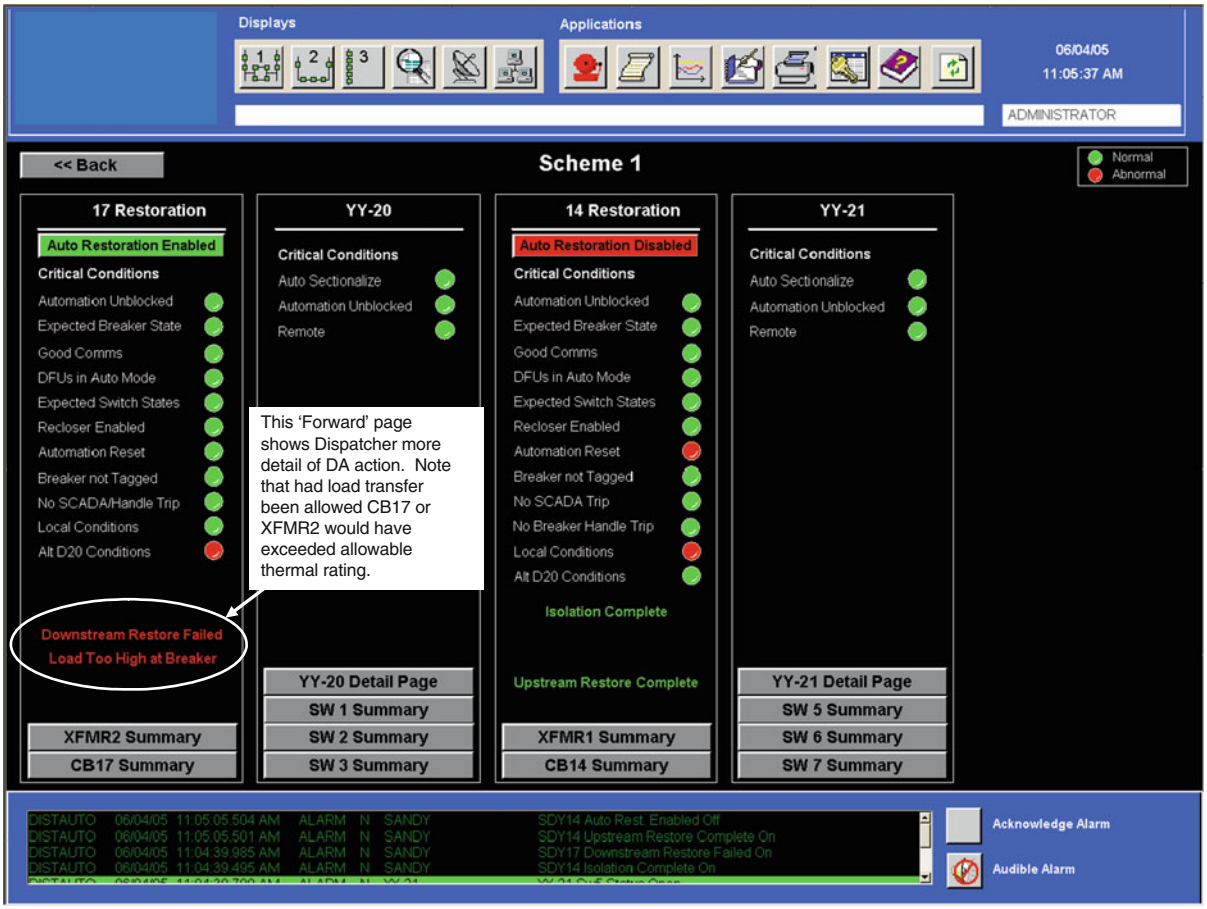
Considerations

Centralized, Decentralized, or Local Algorithm Given the increasing sophistication of various devices in the system, many utilities are facing a choice of location for the various algorithms (Fig. 22). Often it is driven by the unique characteristics of the devices installed or by the various alternatives provided by the automation equipment suppliers.

Table 2 compares the various schemes.

Safety and Work Processes The safety of workers, of the general public, and of equipment must not be compromised. This imposes the biggest challenge for deploying any automatic or remotely controlled systems. New automation systems often require new work processes. Utility work process and personnel must be well trained to safely operate and maintain the new automated distribution grid systems.

Operating practices and procedures must be reviewed and modified as necessary to address the presence of automatic switchgear.



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 14
Dispatch detail of scheme 1 restoration failure

Safety related recommendations include:

- Requirement for "visible gap" for disconnect switches
- No automatic closures after 2 min have elapsed following the initial fault to protect line crews
- System disabled during maintenance ("live line") work, typically locally and remotely

The Law of Diminishing Returns Larger utilities serve a range of customer types across a range of geographic densities. Consequently, the voltage profile and the exposure to outages are very different from circuit to circuit. Most utilities analyze distribution circuits and deploy automation on the most troublesome feeders first. Figure 23 depicts this difference.

Figure 23 highlights the decision by one utility to automate roughly 25% of feeders, which account for 70% of overall customer minutes interrupted.

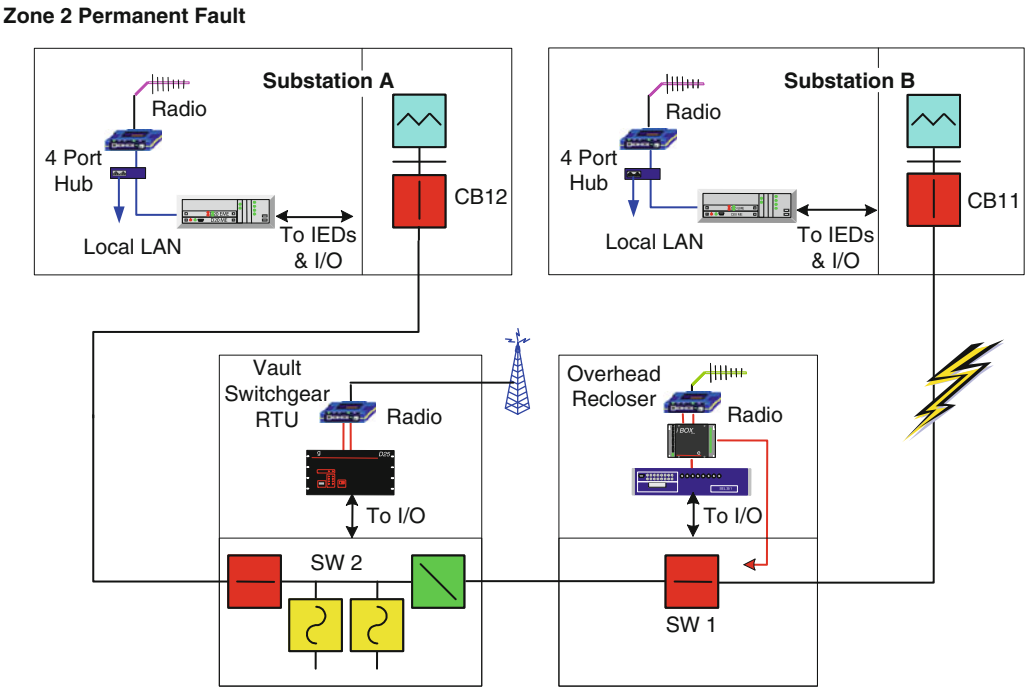
The same analysis can be done on a circuit basis. The addition of each additional sensing and monitoring device to a feeder leads to a diminishing improvement to outage minutes as shown in Fig. 24.

Both of these elements are typically studied and modeled to determine the recommended amount of automation each utility is planning.

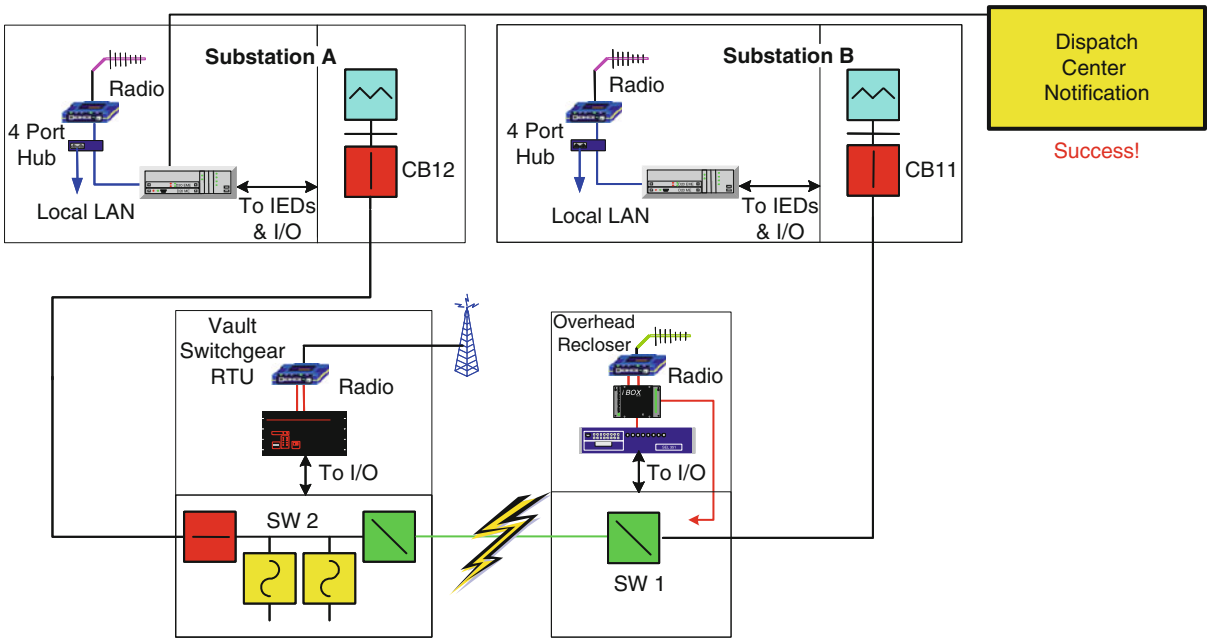
Substations

Role and Types of Substations

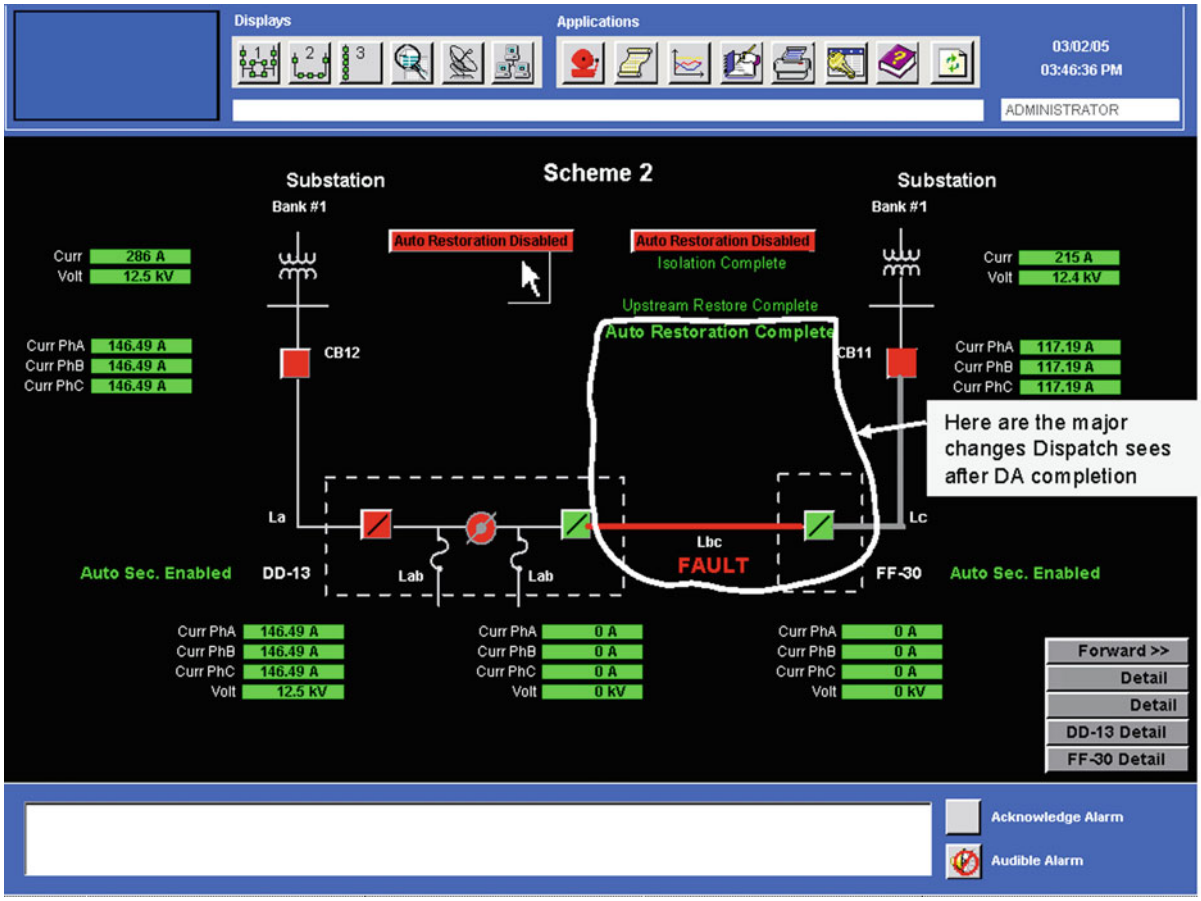
Substations are key parts of electrical generation, transmission, and distribution systems. Substations



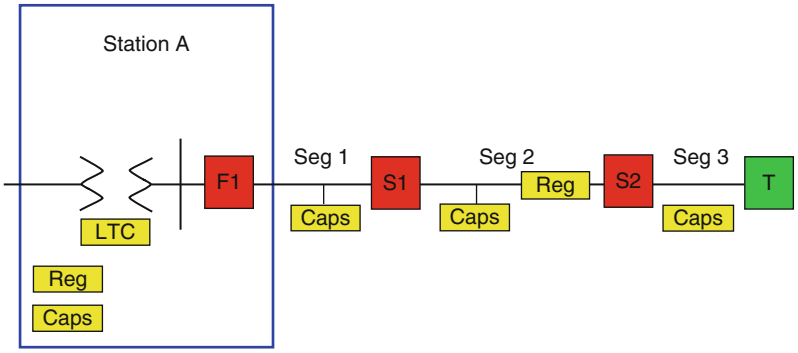
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 15
Scheme 2 architecture



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 16
Scheme 2 architecture after successful DA operation



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 17
Dispatch notification of scheme 2 isolation/restoration success



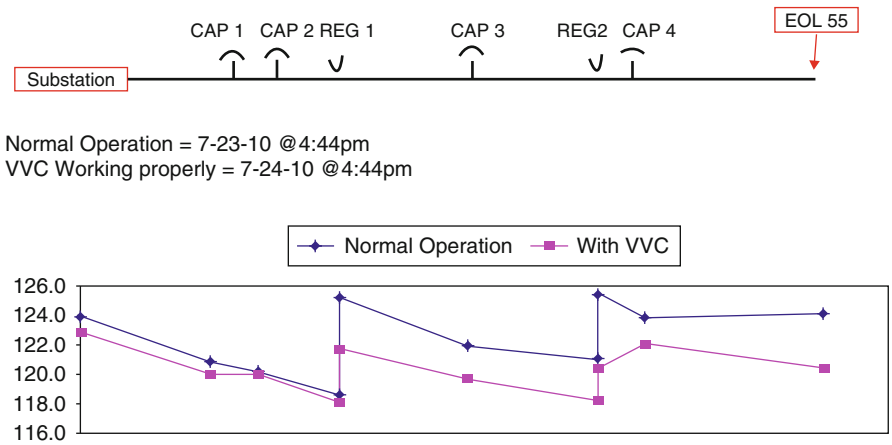
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 18
Example station and feeder voltage/VAR control devices

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 19
Example station and feeder voltage/VAR control devices

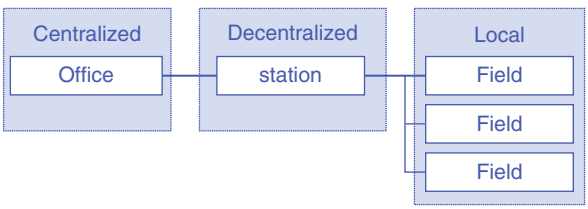
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 20
Three-phase station voltage regulator

transform voltage from high to low or from low to high as necessary. Substations also dispatch electric power from generating stations to consumption centers. Electric power may flow through several substations between the generating plant and the consumer, and the voltage may be changed in several steps. Substations can be generally divided into three major types:

1. *Transmission substations* integrate the transmission lines into a network with multiple parallel interconnections so that power can flow freely over long distances from any generator to any consumer. This transmission grid is often called the bulk power system. Typically, transmission lines operate at voltages above 138 kV. Transmission substations often include transformation from one transmission voltage level to another.
2. *Sub-transmission substations* typically operate at 34.5 kV through 138 kV voltage levels, and transform the high voltages used for efficient



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 21
Three-phase station voltage profile



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 22
Relationship automation at centralized, decentralized, and local areas

long distance transmission through the grid to the sub-transmission voltage levels for more cost-effective transmission of power through supply lines to the distribution substations in the surrounding regions. These supply lines are radial express feeders, each connecting the substation to a small number of distribution substations.

3. *Distribution substations* typically operate at 2.4–34.5 kV voltage levels, and deliver electric energy directly to industrial and residential consumers. Distribution feeders transport power from the distribution substations to the end consumers' premises. These feeders serve a large number of premises and usually contain many branches. At the consumers' premises, distribution transformers transform the distribution voltage to the service level voltage directly used in households and industrial plants, usually from 110 to 600 V.

Recently, distributed generation has started to play a larger role in the distribution system supply. These are small-scale power generation technologies (typically in the range of 3–10,000 kW) used to provide an alternative to or an enhancement of the traditional electric power system. Distributed generation includes combined heat and power (CHP), fuel cells, micro-combined heat and power (micro-CHP), micro-turbines, photovoltaic (PV) systems, reciprocating engines, small wind power systems, and Stirling engines, as well as renewable energy sources.

Renewable energy comes from natural resources such as sunlight, wind, rain, tides, and geothermal heat, which are naturally replenished. New renewables (small hydro, modern biomass, wind, solar, geothermal, and biofuels) are growing very rapidly.

A simplified one-line diagram showing all major electrical components from generation to a customer's service is shown in Fig. 25.

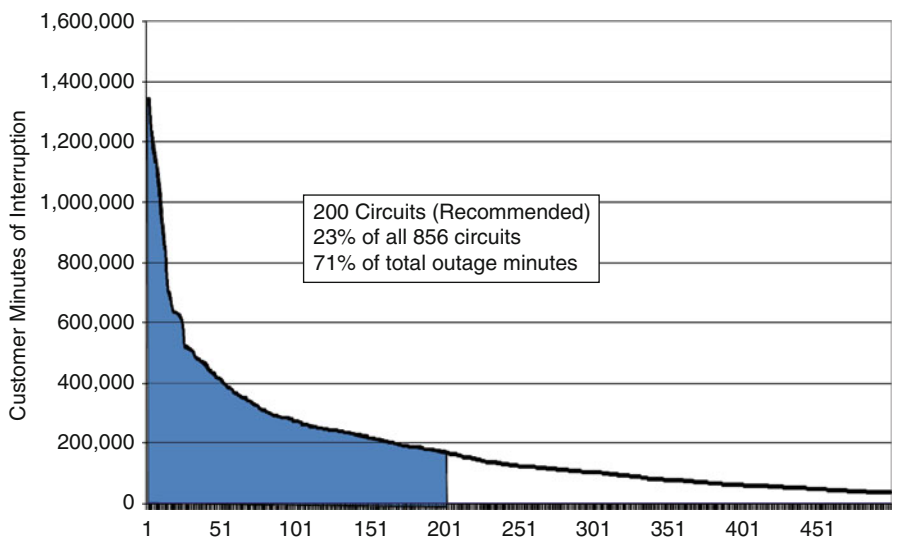
Distribution Substation Components

Distribution substations are comprised of the following major components.

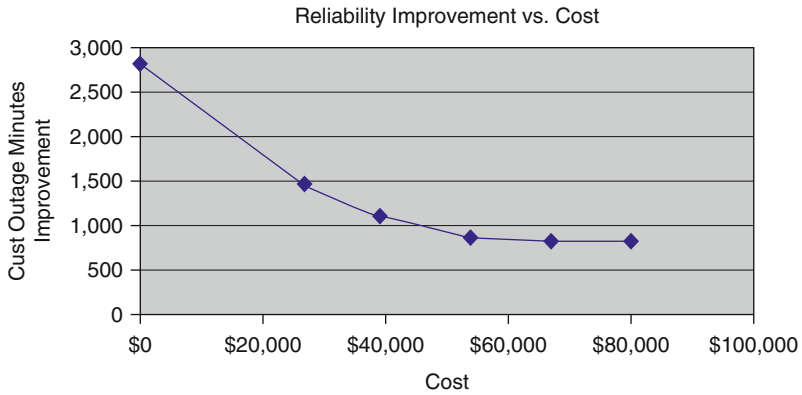
Supply Line Distribution substations are connected to a sub-transmission system via at least one supply line, which is often called a primary feeder. However, it is typical for a distribution substation to be supplied by two or more supply lines to increase reliability of the

Distribution Systems, Substations, and Integration of Distributed Generation. Table 2 Three-phase station voltage profile

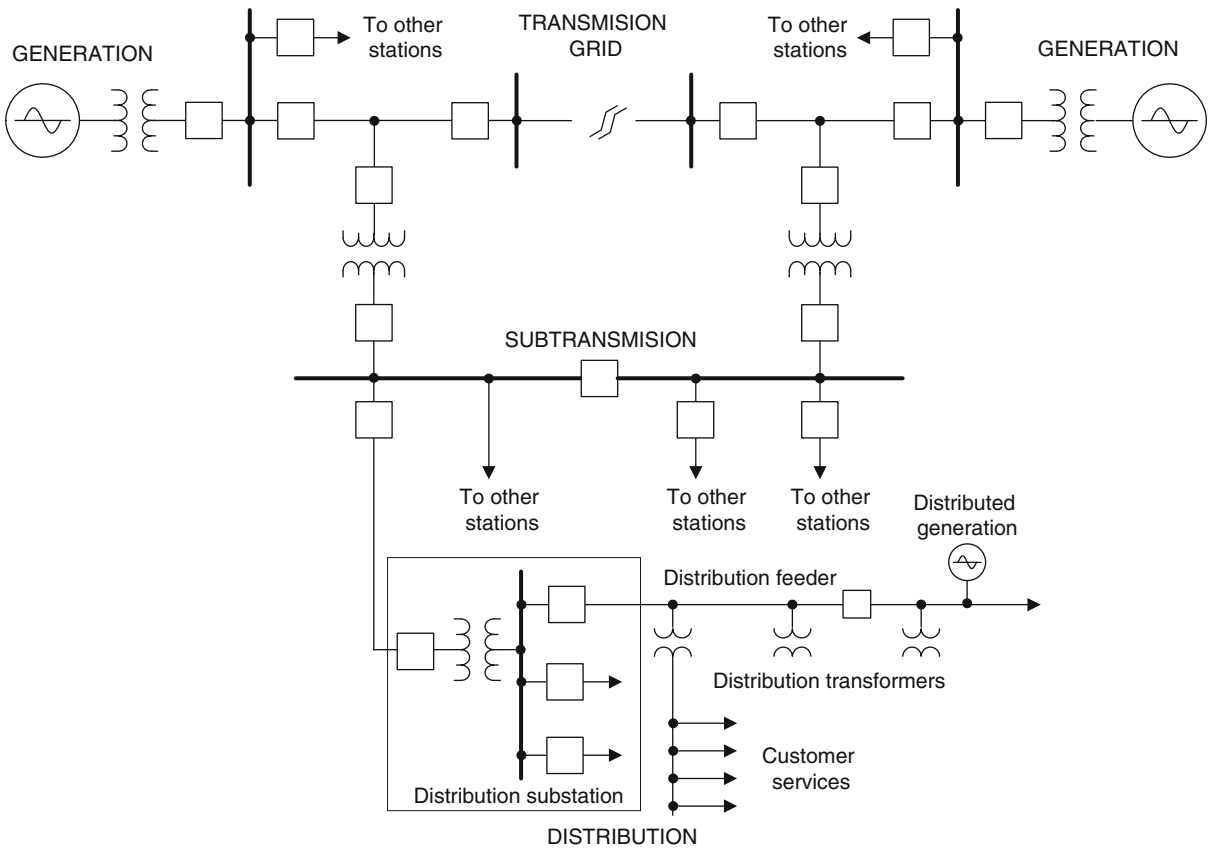
Centralized	Decentralized	Local
Supports more complex applications such as: Load Flow, DTS, Study	Most station IEDs support automation	Local IEDs often include local algorithms
Support for full network model	Faster response than centralized DA	Usually initiated after prolonged comms outage, e.g., local capacitor controller
Optimizes improvements	Smaller incremental deployment costs	Operates faster than other algorithms usually for protection, reclosing, and initial sectionalizing
Dynamic system configuration	Often used for initial deployment because of the reduced complexity and costs	Usually only operates based on local sensing or peer communications
Automation during abnormal conditions	Typical applications: include: initial response, measure pre-event	Less sophisticated and less expensive
Enables integration with other sources of data – EMS, OMS, AMI, GIS	Flexible, targeted, or custom solution	Easiest to begin deploying
Integration with other processes planning, design, dispatch	Usually cheaper and easier for initial deploy	Hardest to scale sophisticated solutions
Easier to scale, maintain, upgrade, and backup	Hard to scale sophisticated solutions	



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 23
Customer minutes interrupted by feeder



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 24
Customer minutes interrupted by cost



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 25
One-line diagram of major components of power system from generation to consumption

power supply in case one supply line is disconnected. A supply line can be an overhead line or an underground feeder, depending on the location of the substation, with underground cable lines mostly in urban areas and overhead lines in rural areas and suburbs. Supply lines are connected to the substation via high-voltage disconnecting switches in order to isolate lines from substation to perform maintenance or repair work.

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 26
Voltage transformers (Courtesy of General Electric)

Transformers Transformers "step down" supply line voltage to distribution level voltage. See Fig. 26. Distribution substations usually employ three-phase transformers; however, banks of single-phase transformers can also be used. For reliability and maintenance purposes, two transformers are typically employed at the substation, but the number can vary depending on the importance of the consumers fed from the substation and the distribution system design in general. Transformers can be classified by the following factors:

(a) *Power rating*, which is expressed in kilovolt-amperes (kVA) or megavolts-amperes (MVA), and indicates the amount of power that can be transferred through the transformer. Distribution substation transformers are typically in the range of 3 kVA to 25 MVA.

(b) *Insulation*, which includes liquid or dry types of transformer insulation. Liquid insulation can be mineral oil, nonflammable or low-flammable liquids. The dry type includes the ventilated, cast coil, enclosed non-ventilated, and sealed gas-filled types. Additionally, insulation can be a combination of the liquid-, vapor-, and gas-filled unit.

(c) *Voltage rating*, which is governed by the sub-transmission and distribution voltage levels substation to which the transformer is connected. Also, there are standard voltages nominal levels governed by applicable standards. Transformer voltage rating is indicated by the manufacturer. For example, 115/34.5 kV means the high-voltage winding of the transformer is rated at 115 kV, and the low-voltage winding is rated at 34.5 kV between different phases. Voltage rating dictates the construction and insulation requirements of the transformer to withstand rated voltage or higher voltages during system operation.

(d) *Cooling*, which is dictated by the transformer power rating and maximum allowable temperature rise at the expected peak demand. Transformer rating includes self-cooled rating at the specified temperature rise or forced-cooled rating of the transformer if so equipped. Typical transformer rated winding temperature rise is 55°C/65°C at ambient temperature of 30°C for liquid-filled transformers to permit 100% loading or higher if temporarily needed for system operation. Modern low-loss transformers allow even higher temperature rise; however, operating at higher temperatures may impact insulation and reduce transformer life.

(e) *Winding connections*, which indicates how the three phases of transformer windings are connected together at each side. There are two basic connections of transformer windings; delta (where the end of each phase winding is connected to the beginning of the next phase forming a triangle); and star (where the ends of each phase winding are connected together, forming a neutral point and the beginning of windings are connected outside). Typically, distribution transformer is connected delta at the high-voltage side and wye at the low-voltage side. Delta connection isolates the two systems with respect to some harmonics (especially third harmonic), which are not desirable in the system. A wye connection establishes a convenient neutral point for connection to the ground.

(f) *Voltage regulation*, which indicates that the transformer is capable of changing the low-voltage side voltage in order to maintain nominal voltage at customer service points. Voltage at customer

service points can fluctuate as a result of either primary system voltage fluctuation or excessive voltage drop due to the high load current. To achieve this, transformers are equipped with voltage tap regulators. Those can be either no-load type, requiring disconnecting the load to change the tap, or under-load type, allowing tap changing during transformer normal load conditions. Transformer taps effectively change the transformation ratio and allow voltage regulation of ±10–15% in steps of 1.75–2.5% per tap. Transformer tap changing can be manual or automatic; however, only under-load type tap changers can operate automatically.

Busbars Busbars (also called buses) can be found throughout the entire power system, from generation to industrial plants to electrical distribution boards. Busbars are used to carry large current and to distribute current to multiple circuits within switchgear or equipment (Fig. 27). Plug-in devices with circuit breakers or fusible switches may be installed and wired without de-energizing the busbars if so specified by the manufacturer.

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 27
Outdoor switchgear busbar (upper conductors) with voltage transformers (Courtesy of General Electric)

Originally, busbars consisted of uncovered copper conductors supported on insulators, such as porcelain, mounted within a non-ventilated steel housing. This type of construction was adequate for current ratings of 225–600 A. As the use of busbars expanded and increased, loads demanded higher current ratings and housings were ventilated to provide better cooling at higher capacities. The busbars were also covered with insulation for safety and to permit closer spacing of bars of opposite polarity in order to achieve lower reactance and voltage drop.

By utilizing conduction, current densities are achieved for totally enclosed busbars that are comparable to those previously attained with ventilated busbars. Totally enclosed busbars have the same current rating regardless of mounting position. Bus configuration may be a stack of one busbar per phase (0–800 A), whereas higher ratings will use two (3,000 A) or three stacks (5,000 A). Each stack may contain all three phases, neutral, and grounding conductors to minimize circuit reactance.

Busbars' conductors and current-carrying parts can be either copper, aluminum, or copper alloy rated for the purpose. Compared to copper, electrical grade aluminum has lower conductivity and lower mechanical strength. Generally, for equal current-carrying ability, aluminum is lighter in weight and less costly. All contact locations on current-carrying parts are plated with tin or silver to prevent oxides or insulating film from building up on the surfaces.

In distribution substations, busbars are used at both high side and low side voltages to connect different circuits and to transfer power from the power supply to multiple outcoming feeders. Feeder busbars are available for indoor and outdoor construction. Outdoor busbars are designed to operate reliably despite exposure to the weather. Available current ratings range from 600 to 5,000 A continuous current. Available short-circuit current ratings are 42,000–200,000 A, symmetrical root mean square (RMS).

Switchgear Switchgear (Fig. 28) is a general term covering primary switching and interrupting devices together with its control and regulating equipment. Power switchgear includes breakers, disconnect switches, main bus conductors, interconnecting

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 28
Indoor switchgear front view (Courtesy of General Electric)

wiring, support structures with insulators, enclosures, and secondary devices for monitoring and control. Power switchgear is used throughout the entire power system, from generation to industrial plants to connect incoming power supply and distribute power to consumers. Switchgear can be of outdoor or indoor types, or a combination of both. Outdoor switchgear is typically used for voltages above 26 kV, whereas indoor switchgear is commonly for voltages below 26 kV.

Indoor switchgear can be further classified into metal-enclosed switchgear and open switchgear, which is similar to outdoor switchgear but operates at lower voltages. Metal-enclosed switchgear can be further classified into metal-clad switchgear, low-voltage breaker switchgear, and interrupter switchgear. Metal-clad switchgear is commonly used throughout the industry for distributing supply voltage service above 1,000 V.

Metal-clad switchgear can be characterized as follows:

(a) The primary voltage breakers and switches are mounted on a removable mechanism to allow for movement and proper alignment.
(b) Grounded metal barriers enclose major parts of the primary circuit, such as breakers or switches, buses, potential transformers, and control power transformers.
(c) All live parts are enclosed within grounded metal compartments. Primary circuit elements are not exposed even when the removable element is in the test, disconnected, or in the fully withdrawn position.
(d) Primary bus conductors and connections are covered with insulating material throughout by means of insulated barriers between phases and between phase and ground.
(e) Mechanical and electrical interlocking ensures proper and safe operation.
(f) Grounded metal barriers isolate all primary circuit elements from meters, protective relays, secondary control devices, and wiring. Secondary control devices are mounted of the front panel, and are usually swing type as shown in Fig. 28.

Switchgear ratings indicate specific operating conditions, such as ambient temperature, altitude, frequency, short-circuit current withstand and duration, overvoltage withstand and duration, etc. The rated continuous current of a switchgear assembly is the maximum current in RMS (root mean square) amperes, at rated frequency, that can be carried continuously by the primary circuit components without causing temperatures in excess of the limits specified by applicable standards.

Outcoming Feeders A number of outcoming feeders are connected to the substation bus to carry power from the substation to points of service. Feeders can be run overhead along streets, or beneath streets, and carry power to distribution transformers at or near consumer premises. The feeders' breaker and isolator are part of the substation low-voltage switchgear and are typically the metal-clad type. When a fault occurs on the feeder, the protection will detect it and open the breaker. After detection, either automatically or manually, there may be one or more attempts to reenergize the feeder. If the fault is transient, the feeder will be reenergized and the breaker will remain closed. If the fault is permanent, the breaker will remain open and operating personnel will locate and isolate the faulted section of the feeder.

Switching Apparatus Switching apparatus is needed to connect or disconnect elements of the power system to or from other elements of the system. Switching apparatus includes switches, fuses, circuit breakers, and service protectors.

(a) *Switches* are used for isolation, load interruption, and transferring service between different sources of supply.

Isolating switches are used to provide visible disconnect to enable safe access to the isolated equipment. These switches usually have no interrupting current rating, meaning that the circuit must be opened by other means (such as breakers). Interlocking is generally provided to prevent operation when the switch is carrying current.

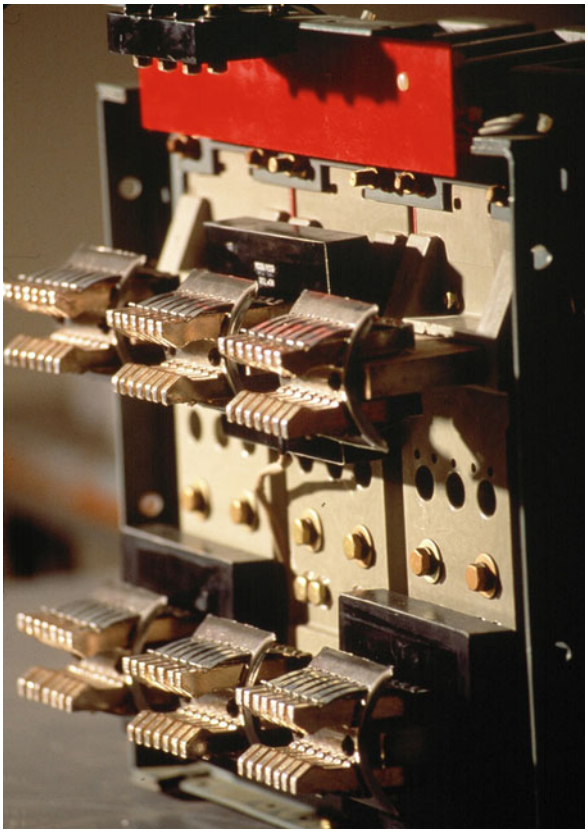
Load interrupting or a load-break switch combines the functions of a disconnecting switch and a load interrupter for interrupting at rated voltage and currents not exceeding the continuous-current rating of the switch. Load-break switches are of the air- or fluid-immersed type. The interrupter switch is usually manually operated and has a "quick-make, quick-break" mechanism which functions independently of the speed-of-handle operation. These types of switches are typically used on voltages above 600 V.

For services of 600 V and below, safety circuit breakers and switches are commonly used. Safety switches are enclosed and may be fused or un-fused. This type of switch is operated by a handle outside the enclosure and is interlocked so that the enclosure cannot be opened unless the switch is open or the interlock defeater is operated.

Transfer switches can be operated automatically or manually. Automatic transfer switches are of double-throw construction and are primarily used for emergency and standby power generation systems rated at 600 V and lower. These switches are used to provide protection against normal service failures.

(b) *Fuses* are used as an over-current-protective device with a circuit-opening fusible link that is heated and severed as over-current passes through it. Fuses are available in a wide range of voltage, current, and interrupting ratings, current-limiting types, and for indoor and outdoor applications. Fuses perform the same function as circuit breakers, and there is no general rule for using one versus the other. The decision to use a fuse or circuit breaker is usually based on the particular application, and factors such as the current interrupting requirement, coordination with adjacent protection devices, space requirements, capital and maintenance costs, automatic switching, etc.

(c) *Circuit breakers* (Fig. 29) are devices designed to open and close a circuit either automatically or manually. When applied within its rating, an automatic circuit breaker must be capable of opening a circuit automatically on a predetermined overload of current without damaging itself or adjacent elements. Circuit breakers are required to operate infrequently, although some classes of circuit breakers are suitable for more frequent operation. The interrupting and momentary ratings of a circuit breaker must be equal to or greater than the available system short-circuit currents.



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 29
Breaker of indoor switchgear, rear "bus" side (Courtesy of General Electric)

Circuit breakers are available for the entire system voltage range, and may be as furnished single-, double-, triple-, or four-pole, and arranged for indoor or outside use. Sulfur hexafluoride (SF_6) gas-insulated circuit breakers are available for medium and high voltages, such as gas-insulated substations.

When a current is interrupted, an arc is generated. This arc must be contained, cooled, and extinguished in a controlled way so that the gap between the contacts can again withstand the voltage in the circuit. Circuit breakers can use vacuum, air, insulating gas, or oil as the medium in which the arc forms. Different techniques are used to extinguish the arc, including:

- Lengthening the arc
- Intensive cooling (in jet chambers)
- Division into partial arcs
- Zero point quenching (contacts open at the zero current time crossing of the AC waveform, effectively breaking no-load current at the time of opening)
- Connecting capacitors in parallel with contacts in DC circuits

Traditionally, oil circuit breakers (Fig. 30) were used in the power industry, which use oil as a media to extinguish the arc and rely upon vaporization of some of the oil to blast a jet of oil through the arc.

Gas (usually sulfur hexafluoride) circuit breakers sometimes stretch the arc using a magnetic field, and then rely upon the dielectric strength of the sulfur hexafluoride to quench the stretched arc.

Vacuum circuit breakers have minimal arcing (as there is nothing to ionize other than the contact material), so the arc quenches when it is stretched by a very small amount (<2–3 mm). Vacuum circuit breakers are frequently used in modern medium-voltage switchgear up to 35 kV.

Air blast circuit breakers may use compressed air to blow out the arc, or alternatively, the contacts are rapidly swung into a small sealed chamber, where the escaping displaced air blows out the arc.

Circuit breakers are usually able to terminate all current very quickly: Typically the arc is extinguished between 30 and 150 ms after the mechanism has tripped, depending upon age and construction of the device.

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 30
Outdoor medium-voltage oil-immersed circuit breaker (Courtesy of General Electric)

Indoor circuit breakers are rated to carry 1–3 kA current continuously, and interrupting 8–40 kA short-circuit current at rated voltage.

Surge Voltage Protection Transient overvoltages are due to natural and inherent characteristics of power systems. Overvoltages may be caused by lightning or by a sudden change of system conditions (such as switching operations, faults, load rejection, etc.), or both. Generally, the overvoltage types can be classified as lightning generated and as switching generated. The magnitude of overvoltages can be above maximum permissible levels, and therefore needs to be reduced and protected against to avoid damage to equipment and undesirable system performance.

The occurrence of abnormal applied overvoltage stresses, either short term or sustained steady state, contributes to premature insulation failure. Large amounts of current may be driven through the faulted

channel, producing large amounts of heat. Failure to suppress overvoltage quickly and effectively or interrupt high short-circuit current can cause massive damage of insulation in large parts of the power system, leading to lengthy repairs.

The appropriate application of surge-protective devices will lessen the magnitude and duration of voltage surges seen by the protected equipment. The problem is complicated by the fact that insulation failure results from impressed overvoltages, and because of the aggregate duration of repeated instances of overvoltages.

Surge arresters have been used in power systems to protect insulation from overvoltages. Historically, the evolution of surge arrester material technology has produced various arrester designs, starting with the valve-type arrester, which has been used almost exclusively on power system protection for decades. The active element (i.e., valve element) in these arresters is a nonlinear resistor that exhibits relatively high resistance (megaohms) at system operating voltages, and a much lower resistance (ohms) at fast rate-of-rise surge voltages.

In the mid-1970s, arresters with metal-oxide valve elements were introduced. Metal-oxide arresters have valve elements (also of sintered ceramic-like material) of a much greater nonlinearity than silicon carbide arresters, and series gaps are no longer required. The metal-oxide designs offer improved protective characteristics and improvement in various other characteristics compared to silicon carbide designs. As a result, metal-oxide arresters have replaced gapped silicon carbide arresters in new installations.

In the mid-1980s, polymer housings began to replace porcelain housings on metal-oxide surge arresters offered by some manufacturers. The polymer housings are made of either EPDM (ethylene propylene diene monomer [M-class] rubber) or silicone rubber. Distribution arrester housings were first made with polymer, and later expanded into the intermediate and some station class ratings. Polymer housing material reduces the risk of injuries and equipment damage due to surge arrester failures.

Arresters have a dual fundamental-frequency (RMS) voltage rating (i.e., duty-cycle voltage rating), and a corresponding maximum continuous operating voltage rating. Duty-cycle voltage is defined as the designated maximum permissible voltage between the terminals at which an arrester is designed to perform.

Grounding Grounding is divided into two categories: power system grounding and equipment grounding.

Power system grounding means that at some location in the system there are intentional electric connections between the electric system phase conductors and ground (earth). System grounding is needed to control overvoltages and to provide a path for ground-current flow in order to facilitate sensitive ground-fault protection based on detection of ground-current flow. System grounding can be as follows:

- Solidly grounded
- Ungrounded
- Resistance grounded

Each grounding arrangement has advantages and disadvantages, with choices driven by local and global standards and practices, and engineering judgment.

Solidly grounded systems are arranged such that circuit protective devices will detect a faulted circuit and isolate it from the system regardless of the type of fault. All transmission and most sub-transmission systems are solidly grounded for system stability purposes. Low-voltage service levels of 120–480 V four-wire systems must also be solidly grounded for safety of life. Solid grounding is achieved by connecting the neutral of the wye-connected winding of the power transformer to the ground.

Where service continuity is required, such as for a continuously operating process, the resistance grounded power system can be used. With this type of grounding, the intention is that any contact between one phase conductor and a ground will not cause the phase over-current protective device to operate. Resistance grounding is typically used from 480 V to 15 kV for three-wire systems. Resistance grounding is achieved by connecting the neutral of the wye-connected winding of the power transformer to the ground through the resistor, or by employing special grounding transformers.

The operating advantage of an ungrounded system is the ability to continue operations during a single phase-to-ground fault, which, if sustained, will not

result in an automatic trip of the circuit by protection. Ungrounded systems are usually employed at the distribution level and are originated from delta-connected power transformers.

Equipment grounding refers to the system of electric conductors (grounding conductor and ground buses) by which all non-current-carrying metallic structures within an industrial plant are interconnected and grounded. The main purposes of equipment grounding are:

- To maintain low potential difference between metallic structures or parts, minimizing the possibility of electric shocks to personnel in the area
- To contribute to adequate protective device performance of the electric system, and safety of personnel and equipment
- To avoid fires from volatile materials and the ignition of gases in combustible atmospheres by providing an effective electric conductor system for the flow of ground-fault currents and lightning and static discharges to eliminate arcing and other thermal distress in electrical equipment

Substation grounding systems are thoroughly engineered. In an electrical substation, a ground (earth) mat is a mesh of metal rods connected together with conductive material and installed beneath the earth surface. It is designed to prevent dangerous ground potential from rising at a place where personnel would be located when operating switches or other apparatus. It is bonded to the local supporting metal structure and to the switchgear so that the operator will not be exposed to a high differential voltage due to a fault in the substation.

Power Supply Quality The quality of electrical power may be described as a set of values or parameters, such as:

- Continuity of service
- Variation in voltage magnitude
- Transient voltages and currents
- Harmonic content in the supply voltages

Continuity of service is achieved by proper design, timely maintenance of equipment, reliability of all substation components, and proper operating procedures. Recently, remote monitoring and control have greatly improved the power supply continuity.

When the voltage at the terminals of utilization equipment deviates from the value on the nameplate of the equipment, the performance and the operating life of the equipment are affected. Some pieces of equipment are very sensitive to voltage variations (e.g., motors). Due to voltage drop down the supply line, voltage at the service point may be much lower compared with the voltage at substation. Abnormally low voltage occurs at the end of long circuits. Abnormally high voltage occurs at the beginning of circuits close to the source of supply, especially under lightly loaded conditions such as at night and during weekends. Voltage regulators are used at substations to improve the voltage level supplied from the distribution station. This is achieved by a tap changer mounted in the transformer and an automatic voltage regulator that senses voltage and voltage drop due to load current to increase or decrease voltage at the substation.

If the load power factor is low, capacitor banks (Fig. 31) may be installed at the substation to improve the power factor and reduce voltage drop. Capacitor banks are especially beneficial at substations near industrial customers where reactive power is needed for operation of motors.

Transients in voltages and currents may be caused by several factors, such as large motor stating, fault in the sub-transmission or distribution system, lightning, welding equipment and arc furnace operation, turning on or off large loads, etc. Lighting equipment output is sensitive to applied voltage, and people are sensitive to sudden illumination changes. A voltage change of 0.25–0.5% will cause a noticeable reduction in the light output of an incandescent lamp. Events causing such voltage effects are called flicker (fast change of the supply voltage), and voltage sags (depressed voltage for a noticeable time). Both flicker and sags have operational limits and are governed by industry and local standards.

Voltage and current on the ideal AC power system have pure single frequency sine wave shapes. Power systems have some distortion because an increasing number of loads require current that is not a pure sine wave. Single- and three-phase rectifiers, adjustable speed drives, arc furnaces, computers, and fluorescent lights are good examples. Capacitor failure, premature transformer failure, neutral overloads, excessive motor

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 31
Capacitor Bank (Courtesy of General Electric)

heating, relay misoperation, and other problems are possible when harmonics are not properly controlled.

Harmonics content is governed by appropriate industry and local standards, which also provide recommendations for control of harmonics in power systems.

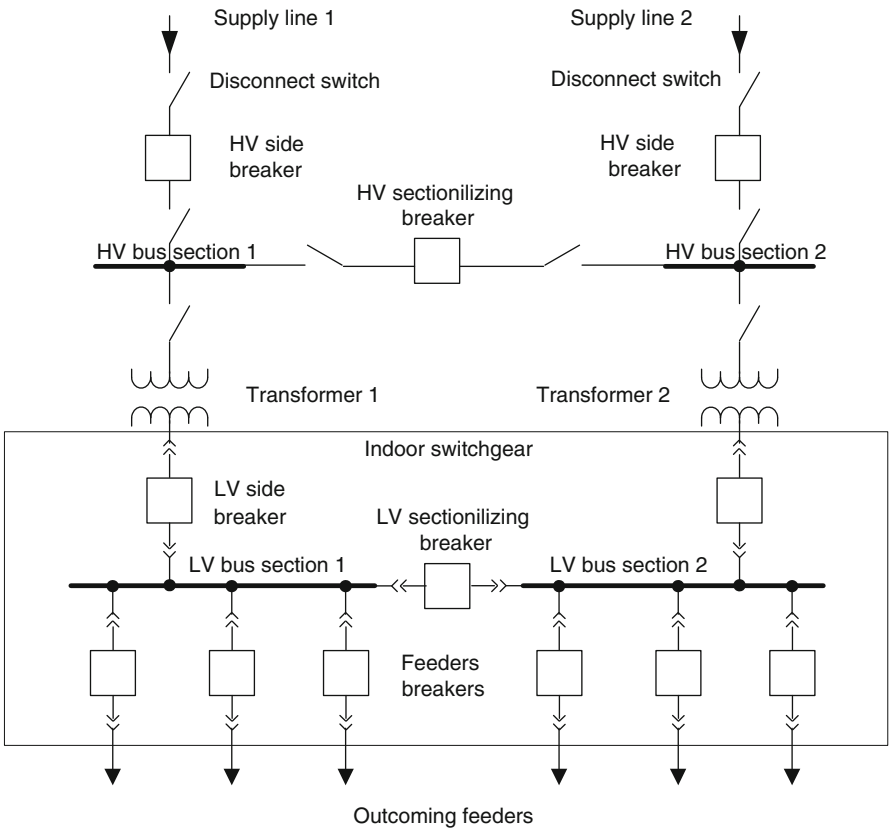
Substation Design Considerations

Distribution substation design is a combination of reliability and quality of the power supply, safety, economics, maintainability, simplicity of operation, and functionality.

Safety of life and preservation of property are the two most important factors in the design of the substation. Codes must be followed and recommended practices or standards should be followed in the selection and application of material and equipment. Following are the operating and design limits that should be considered in order to provide safe working conditions:

- Interrupting devices must be able to function safely and properly under the most severe duty to which they may be exposed.
- Accidental contact with energized conductors should be eliminated by means of enclosing the conductors, installing protective barriers, and interlocking.
- The substation should be designed so that maintenance work on circuits and equipment can be accomplished with these circuits and equipment de-energized and grounded.
- Warning signs should be installed on electric equipment accessible to both qualified and unqualified personnel, on fences surrounding electric equipment, on access doors to electrical rooms, and on conduits or cables above 600 V in areas that include other equipment.
- An adequate grounding system must be installed.
- Emergency lights should be provided where necessary to protect against sudden lighting failure.
- Operating and maintenance personnel should be provided with complete operating and maintenance instructions, including wiring diagrams, equipment ratings, and protective device settings.

A variety of basic circuit arrangements are available for distribution substations. Selection of the best system or combination of systems will depend upon the needs of the power supply process. In general, system costs increase with system reliability if component quality is equal. Maximum reliability per unit investment can be achieved by using properly applied and well-designed components.



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 32
Example one-line diagram of distribution substation with two transformers and two supply lines

Figure 32 provides an example of the distribution substation one-line diagram with two transformers, two supply lines, and two sections at both the high-voltage (HV) side and low-voltage (LV) sides, with sectionalizing breakers at both HV and LV voltages. Such an arrangement provides redundancy and reliability in case of any component failure by transferring the power supply from one section to another. Additionally, any component of the substation can be taken out of service for maintenance.

If the substation is designed to supply a manufacturing plant, continuity of service may be critical. Some plants can tolerate interruptions while others require the highest degree of service continuity. The system should always be designed to isolate faults with a minimum disturbance to the system, and should have features to provide the maximum dependability consistent with the plant requirements and justifiable cost. The majority of utilities today supply energy to medium and large industrial customers directly at 34.5, 69, 115, 138, 161, and 230 kV using dedicated substations. Small industrial complexes may receive power at voltages as low as 4 kV.

Poor voltage regulation is harmful to the life and operation of electrical equipment. Voltage at the utilization equipment must be maintained within equipment tolerance limits under all load conditions, or equipment must be selected to operate safely and efficiently within the voltage limits. Load-flow studies and motor-starting calculations are used to verify voltage regulation.

Substation Standardization

Standards, recommended practices, and guides are used extensively in communicating requirements for

design, installation, operation, and maintenance of substations. Standards establish specific definitions of electrical terms, methods of measurement and test procedures, and dimensions and ratings of equipment. Recommended practices suggest methods of accomplishing an objective for specific conditions. Guides specify the factors that should be considered in accomplishing a specific objective. All are grouped together as standards documents.

Standards are used to establish a small number agreed to by the substation community of alternative solutions from a range of possible solutions. This allows purchasers to select a specific standard solution knowing that multiple vendors will be prepared to supply that standard, and that different vendor's produces will be able to interoperate with each other. Conversely, this allows vendors to prepare a small number of solutions knowing that a large number of customers will be specifying those solutions. Expensive and trouble-prone custom "one-of-a-kind" design and manufacturing can be avoided. For example, out of the almost infinite range of voltage ratings for 3-wire 60 Hz distribution substation low side equipment, NEMA C84.1 standardizes only seven: 2,400, 4,160, 4,800, 6,900, 13,800, 23,000, and 34,500 V. Considerable experience and expertise goes into the creation and maintenance of standards, providing a high degree of confidence that solutions implemented according to a standard will perform as expected. Standards also allow purchasers to concisely and comprehensively state their requirements, and allow vendors to concisely and comprehensively state their products' performance.

There are several bodies publishing standards relevant to substations. Representative of these are the following:

The Institute of Electrical and Electronics Engineers (IEEE) is a nonprofit, transnational professional association having 38 societies, of which the Power and Energy Society (PES) is "involved in the planning, research, development, construction, installation, and operation of equipment and systems for the safe, reliable, and economic generation, transmission, distribution, measurement, and control of electric energy." PES includes several committees devoted to various aspects of substations that publish a large number of standards applicable to substations. For more information, visit www.ieee-pes.org.

The National Electrical Manufacturers Association (NEMA) is a trade association of the electrical manufacturing industry that manufactures products used in the generation, transmission and distribution, control, and end-use of electricity. NEMA provides a forum for the development of technical standards that are in the best interests of the industry and users; advocacy of industry policies on legislative and regulatory matters; and collection, analysis, and dissemination of industry data. For more information, visit www.nema.org.

American National Standards Institute (ANSI) oversees the creation, promulgation, and use of thousands of norms and guidelines that directly impact businesses in nearly every sector, including energy distribution. ANSI is also actively engaged in accrediting programs that assess conformance to standards – including globally recognized cross-sector programs such as the ISO 9000 (quality) and ISO 14000 (environmental) management systems. For more information, see www.ansi.org.

National Fire Protection Association (NFPA) is an international nonprofit organization established in 1896 to reduce the worldwide burden of fire and other hazards on the quality of life by providing and advocating consensus codes and standards, research, training, and education. NFPA develops, publishes, and disseminates more than 300 consensus codes and standards intended to minimize the possibility and effects of fire and other risks. Of particular interest to substations is the National Electrical Code (NEC). For more information, see www.nfpa.org.

International Electrotechnical Commission (IEC) technical committee is an organization for the preparation and publication of International Standards for all electrical, electronic, and related technologies. These are known collectively as "electrotechnology." IEC provides a platform to companies, industries, and governments for meeting, discussing, and developing the International Standards they require. All IEC International Standards are fully consensus based and represent the needs of key stakeholders of every nation

participating in IEC work. Every member country, no matter how large or small, has one vote and a say in what goes into an IEC International Standard. For more information, see www.iec.ch.

International Organization for Standardization (ISO) is a nongovernmental organization that forms a bridge between the public and private sectors. Many of its member institutes are part of the governmental structure of their countries, or are mandated by their government while other members have their roots uniquely in the private sector, having been set up by national partnerships of industry associations. ISO enables a consensus to be reached on solutions that meet both the requirements of business and the broader needs of society. For more information, see www.iso.org.

International Telecommunication Union (ITU) is the United Nations' agency for information and communication technology issues, and the global focal point for governments and the private sector in developing networks and services. For 145 years, ITU has coordinated the shared global use of the radio spectrum, promoted international cooperation in assigning satellite orbits, worked to improve telecommunication infrastructure in the developing world, established the worldwide standards that foster seamless interconnection of a vast range of communications systems, and addressed the global challenges of our times, such as mitigating climate change and strengthening cyber security. For more information, see www.iut.int.

Substation Physical Appearance

It is desirable to locate distribution substations as near as possible to the load center of its service area, though this requirement is often difficult to satisfy. Locations that are perfect from engineering and cost points of view are sometimes prohibited due to physical, electrical, neighboring, or aesthetic considerations. Given the required high- and low-voltage requirements and the required power capacity, a low-cost aboveground design will satisfy power supply needs. However, an aboveground design requires overhead sub-transmission line structures (Fig. 33), which are often undesirable structures for neighborhoods. Therefore, incoming lines are often underground cables, and the entire substation is designed to be as unobtrusive as possible.

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 33
Distribution substation view (Courtesy of General Electric)

Underground cables are more costly than overhead lines for the same voltage and capacity. Installation of underground cables in urban areas is also often inconvenient and potentially hazardous.

Local conditions or system-wide policy may require landscaping at a substation site or for housing the substation within a building. A fence with warning signs or other protective enclosures is typically provided to keep unauthorized persons from coming in close proximity of the high-voltage lines and substation equipment. Alarm systems and video surveillance are becoming normal practices for monitoring and preventing intrusion by unauthorized persons.

Protection and Automation

The purpose of protection in distribution substations is to isolate faulted power system elements, such as feeders and transformers, from sources of electrical supply in order to:

- Prevent damage to un-faulted equipment that might otherwise result from sustained fault-level currents and/or voltages

- Reduce the probability and degree of harm to the general public, utility personnel, and property
- Reduce the amount of damage the faulted element sustains, thus containing repair costs, service interruption duration, and impact on the environment
- Clear transient faults and restore service

To accomplish this, protection devices must be able to rapidly determine that a fault has occurred, determine which system element or which section of the system has faulted, and to open the circuit breakers and switches that will disconnect the faulted element. To do this reliably, means must be provided to clear faults even in the event of a single failure in the protection system.

Substation automation facilities complement protection at distribution substations. In the past, automation at distribution substations was limited to automatic tap changer control and capacitor auto-switching to regulate voltage. Automation and communication facilities in modern distribution substations provide visibility to the system operator of the state of the substation, allowing rapid identification of the source and cause of interruptions and other troubles, and providing the ability to dispatch repair personnel quickly to the correct location and with the necessary equipment and spares to effect a repair. In these modern distribution substations, automation facilities often provide operators with the ability to remotely open and close breakers and switches, allowing power rerouting to restore service.

The principle by which protection operates depends to a large degree on the element covered by the protection. Common element types are as follows.

Incoming Sub-transmission Supply Line Protection at Distribution Substations In the past, distribution substations supplied only loads; there was little if any distributed generation. With this arrangement, faults on the incoming sub-transmission supply line do not result in energy flowing out of the distribution substation and into the supply line (i.e., backfeed). In that event, opening of the distribution substation end of the supply line is not necessary to isolate supply line faults; opening of the source end by protection located there is all that is necessary.

Recently, distributed generation from windmills, bio-digesters, mini-hydraulics, and photovoltaic sources has been increasing dramatically, in some cases to the point where the contribution to the supply line faults and must be interrupted to extinguish the fault. Various protection technologies exist to detect such faults, including reverse flow, distance, and current differential.

Reverse flow protection detects the reversal from the normal direction of power and/or current flow. This type of protection is only suitable for cases where the amount of distributed generation in relation to the amount of load is small enough that under normal un-faulted conditions the flow from the supply line is always into the distribution substation. Reverse flow occurs when a supply line fault causes voltage depression to the point where load current essentially disappears, yet the distributed generators continue to source current.

Distance protection estimates the impedance of the supply line between the distribution substation and a fault on the supply line. Supply lines typically have a constant amount of impedance per unit distance (e.g., per kilometer), so the impedance seen at the distribution substation is larger for a fault beyond the end of the supply line than for a fault on the supply line. Distance protection thus can usually discriminate between these two faults, and avoid unnecessary tripping for faults beyond the supply line. To a first approximation, impedance is simply the voltage divided by the current.

Current differential protection is based on the fact that the algebraic sum of all supply line terminal currents is equal to the supply line fault current, if any. Current resulting from external faults and from load flowing into the supply line is canceled as it flows out. Charging current is typically negligible at sub-transmission voltage levels.

With each of these three methods, current transformers (CTs) are needed to measure supply line current, and except for current differential, voltage transformers (VTs) are needed to measure the voltage at the distribution substation. With current differential, a communications channel to the other end(s) of the supply line is needed. A relay (also known as an IED) is needed to detect the supply line fault and discriminate it from other conditions such as

faults on other elements. Also, a circuit breaker or circuit switcher is needed to interrupt the supply line connection on command from the relay.

Distribution Substation Transformer Protection In distribution substations with relatively small delta/wye transformers (typically less than 10 MVA), transformer protection often consists of simple fusing. A power fault in the transformer or on the low side bus typically produces a large current in the high side terminals that melts the fusible links, interrupting the fault. Internal turn-to-turn winding faults that produce only small terminal fault currents are allowed to burn until the fault evolves into a fault with current large enough to operate the fuses. This is acceptable because there is little or no difference in the commercial value of faulted transformers, and because small transformers on which fuses are used are less costly.

On larger (and more expensive) transformers, through-current restrained differential relays (Fig. 34) are most often employed. These protections require CTs measuring all current paths in and out of the transformer. Most often, the CTs are mounted on the high side terminals of the transformer and on the breaker(s) that directly connect to the low side terminals. Where there is no breaker between the transformer and the low side bus, the bus is also covered. The operating principle of transformer current differential protection is based on the incoming high side current under normal and under external fault conditions approximately equaling the sum of the outgoing low side currents after adjusting for the transformer's transformation ratio and internal connection (e.g., delta/wye). During internal fault conditions, there is a large inequality.

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 34
Modern digital transformer protection relay (Courtesy of General Electric)

Invariably, there is actually some difference between these (even in a healthy transformer) that is a small but constant percentage of the current flowing through the transformer. An on-load tap changer is an example of what can cause such a difference. For large through-currents, such as occurring during external low side faults, the difference is substantial, so the relay employs through-current restraint. The differential operating threshold is made dynamic by including in the trip threshold a percentage of the measured through-current.

Another problem with transformer differential protection is that when the transformer itself is saturated, which usually occurs on energization, relatively large currents flow through the excitation impedance, which appear to the protection as differential current with no through-current. Energization current, however, contains a large harmonic component. Relays therefore monitor for the presence of such harmonics, most often the second harmonic, and when found block the differential element.

When a transformer fault is detected, the transformer is isolated by opening its high side connection. Where there are fuses, the fuses perform the interruption. Where there is a high side breaker or circuit switcher, they are tripped. Otherwise, a trip signal is sent to trip the remote end of the supply line. This signal may be via radio, optical fiber, or metallic pair, or may be the closing a switch that grounds one phase of the supply line causing the remote end to see and trip as for a supply line fault. Where there are distributed generators on the distribution system, the distributed generators or the distribution substation's low side breakers must be opened as well. Opening of low side breakers is often done even when there are no distributed generators.

Distribution Substation Bus Protection When fuses protect a distribution substation transformer and there is no transformer low side breaker and the differential

uses feeder breaker CTs, the transformer protection covers the low side bus as well as the transformer.

Where a separate bus protection is used, there are several alternatives for detecting bus faults, including low impedance current differential, high impedance differential, and zone-interlocked schemes.

Low impedance bus differential with through-current restraint operates on the same basic principle as transformer differential protection. Buses have no significant energization current, so harmonic blocking is not implemented. Also there is no need to adjust for transformer connection. However, there are typically many more feeders to measure than in transformer differential applications, so a relay designed for bus protection is required. Low impedance bus differential protections without through-current restraint have been deployed, but often have trouble with CT saturation, causing unnecessary tripping.

High impedance bus differential protections include a simple over-current relay with a large stabilizing resister connected in series. The relay current is from the parallel connection of the secondary windings of CTs measuring all bus connections. The CTs should be identical in type and ratio. With no fault, the current flowing into the bus is balanced by the current flowing out of the bus (a pattern that is matched in the CT secondary circuit), and thus little or no current flows through the over-current relay. The stabilizing resistor tends to prevent large external fault currents from saturating the CT of the faulted feeder and producing erroneous differential current.

Zone-interlocked schemes (also known as bus-blocking schemes) employ an over-current relay measuring infeed current from the main transformer. On detecting fault current, this over-current relay trips the bus unless one of the feeder protections sees a feeder fault and sends a blocking signal. Bus tripping is delayed by a short time to ensure that the feeder protections have sufficient time to detect and block for feeder faults.

No matter the detection method used, when a bus fault is detected, the fault current coming from the main transformer must be interrupted. Where there is a transformer low side breaker, it is opened. Otherwise, the transformer high side is opened using one or more of the methods described in the transformer protection section. Where there are distributed generators on the distribution system, the distributed generators or the distribution substation's low side breakers or circuit switchers must be opened as well. Opening of low side breakers is often done even when there are no distributed generators.

Distribution Feeder Protection Distribution feeder protection at distribution substations is most often instantaneous and timed over-current functionality. Very often, these functions are implemented within a recloser, which is essentially a circuit breaker packaged with a mechanism that repeatedly trips when current exceeds a set threshold, and then recloses or locks out after a series of set delays. Alternatively, an electronic relay with instantaneous and inverse time over-current features is used.

Distribution feeders (Fig. 35) typically have fuses along its length and/or on lateral sections intended to separate a faulted section, allowing the rest to be

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 35
Modern digital feeder protection relay (Courtesy of General Electric)

supplied pending repair. Feeder protections usually implement either a fuse-saving or a trip-saving scheme.

Aerial feeders are prone to transient faults, faults that due to lightning or winding-induced swaying (the technical term is "galloping") disappear when the fault current is interrupted and the arc extinguished. A fuse-saving scheme is often used on these. Such a scheme has in normal conditions an instantaneous over-current element that trips the feeder breaker immediately upon detecting a fault. High speed tripping is used to prevent any fuses between the substation and the fault from melting. After the fault current has been interrupted, the instantaneous element is blocked, an inverse time element is enabled, and the feeder breaker reclosed. For transient faults, on reclosing, the fault will have disappeared and supply to all customers is restored. After a short delay of typically a few seconds, the scheme is reset and the instantaneous element placed back in service. For permanent faults, on reclosing, fault current will again flow, and the inverse time relay will start to operate. The inverse time relay is set to coordinate with the fuses such that if there is a fuse between the substation and the fault, the fuse will operate, interrupting customers on the section it supplies, but leaving all others energized. If the fault is located before any fuse, the inverse time relay will time out and trip the entire feeder.

Underground cable feeders rarely have transient faults; cable faults are almost always permanent. A trip-saving scheme is often used on these. Such a scheme has in normal conditions an inverse time element that trips the feeder only after sufficient time for any fuse between the substation and the fault to operate, interrupting customers on the section it supplies, but leaving all others energized. If the fault is located before any fuse, the inverse time relay will time out and trip the entire feeder. Following the feeder trip, an automatic reclose with instantaneous tripping may be attempted to restore service in cases of a fuse having melted but having been unable to interrupt fault current prior to the first trip.

Occasionally, distribution substation feeder protection includes distance supervision to prevent unnecessary tripping feeder tripping for faults on the low side of large distribution transformers near the substation. Such faults should instead be cleared at the distribution transformer location so that other customers on the feeder are not interrupted.

High Penetration of Distributed Generation and Its Impact on System Design and Operations

High penetration of distributed generation presents significant challenges to design and engineering practices as well as to the reliable operation of the electrical distribution system. The large-scale implementation of distributed energy resources (DER) on system design, performance, and reliable operation requires an integrated approach focused on interoperability, adaptability, and scalability.

Vision for Modern Utilities

Centralized Versus Distributed Generation The bulk of electric power used worldwide is produced at central power plants, most of which utilize large fossil fuel combustion, hydro or nuclear reactors. A majority of these central stations have an output between 30 MW (industrial plant) and 1,700 MW. This makes them relatively large in terms of both physical size and facility requirements as compared with DG alternatives. In contrast, DG is:

- Installed at various locations (closer to the load) throughout the power system
- Not centrally dispatched (although the development of "virtual" power plants, where many decentralized DG units operate as one single unit, may be an exception to this definition)
- Defined by power rating in a wide range from a few kW to tens of MW (in some countries MW limitation is defined by standards, e.g., US, IEEE 1547 defines DG up to 10 MW – either as a single unit or aggregate capacity)
- Connected to the distribution/medium-voltage network – which generally refers to the part of the network that has an operating voltage of 600 V up to 110 kV (depends on the utility/country)

The ownership of the DG is not a factor as to whether a power generator is classified as DG. DG can be owned or operated by electric customers, energy service companies, independent power producers (IPP), or utilities.

The main reasons why central, rather than distributed, generation still dominates current electricity production include economy of scale, fuel cost and availability, and lifetime. Increasing the size of a production unit decreases the cost per MW; however, the advantage of economy of scale is decreasing – technological advances in fuel conversion have improved the economy of small units. Fuel cost and availability is still another reason to keep building large power plants. Additionally, with a lifetime of 25–50 years, large power plants will continue to remain the prime source of electricity for many years to come [1].

The benefits of distributed generation include: higher efficiency; improved security of supply; improved demand-response capabilities; avoidance of overcapacity; better peak load management; reduction of grid losses; network infrastructure cost deferral (CAPEX deferral); power quality support; improved reliability; and environmental and aesthetic concerns (offers a wide range of alternatives to traditional power system design). DG offers extraordinary value because it provides a flexible range of combinations between cost and reliability. In addition, DG may eventually become a more desirable generation asset because it is "closer" to the customer and is more economical than central station generation and its associated transmission infrastructure [2]. The disadvantages of DG are ownership and operation, fuel delivery (machine-based DG, remote locations), cost of connection, dispatchability, and controllability (wind and solar).

Development of "Smart Grid"

In recent years, there has been rapidly growing interest in what is called "smart grid – digitized grid – grid of the future." The concept of smart grids has many definitions and interpretations dependent on the specific country, region, and industry stakeholder's drivers and desirable outcomes and benefits.

The Smart Grids European Technology Platform (which is comprised of European stakeholders, including the research community) defines "a Smart Grid [as] an electricity network that can intelligently integrate the actions of all users connected to it – generators, consumers and those that do both, in order to efficiently deliver sustainable, economic and secure electricity supply" [1].

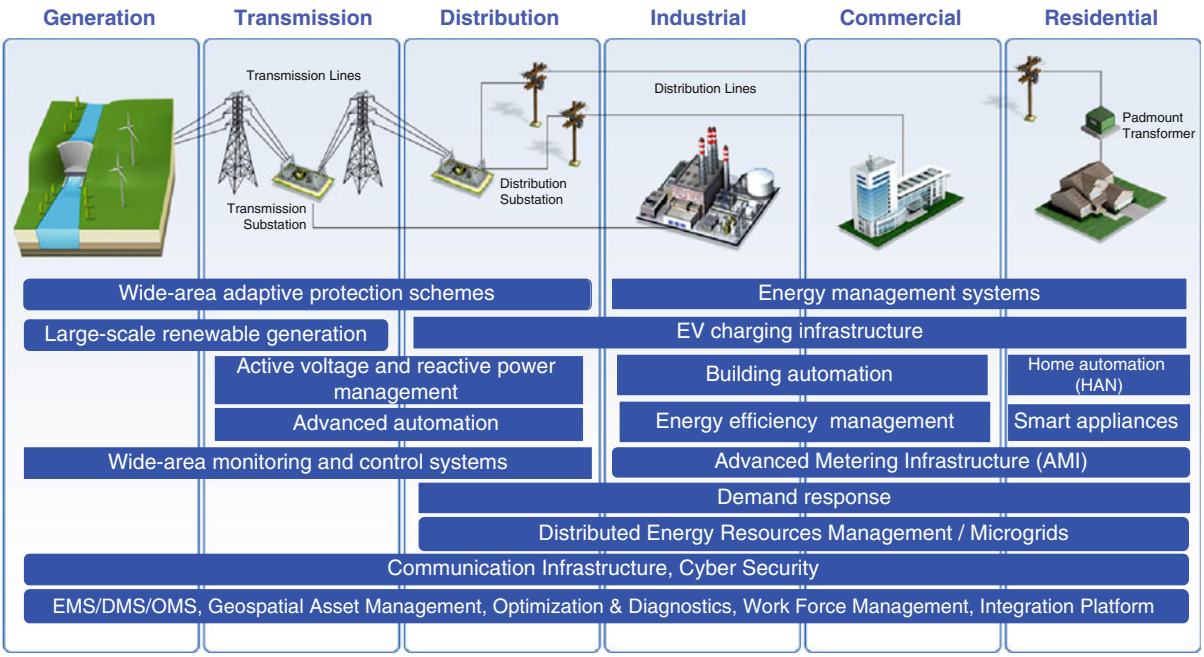
In North America, the two dominant definitions of the smart grid come from the Department of Energy (DOE) and the Electric Power Research Institute (EPRI).

- *US DOE:* "Grid 2030 envisions a fully automated power delivery network that monitors and controls every customer and node, ensuring two-way flow of information and electricity between the power plant and the appliance, and all points in between" [2].
- *EPRI:* "The term 'Smart Grid' refers to a modernization of the electricity delivery system so it monitors, protects, and automatically optimizes the operation of its interconnected elements — from the central and distributed generator through the high-voltage network and distribution system, to industrial users and building automation systems, to energy storage installations and to end-use consumers and their thermostats, electric vehicles, appliances, and other household devices" [3].

Beyond a specific, stakeholder-driven definition, smart grids should refer to the entire power grid from generation through transmission and distribution infrastructure all the way to a wide array of electricity consumers (Fig. 36).

Effective deployment of smart grid technologies requires well-defined and quantified benefits. These benefits can be quantified in the areas of technical and business performance, environmental goals, security of electricity supply, and macro-economic growth and business sustainability development. One of the key components to effectively enable full-value realization is technology – the wide range of technical functionalities and capabilities deployed and integrated as one cohesive end-to-end solution supported by an approach focused on scalability, interoperability, and adaptability. Smart grid technologies can be broadly captured under the following areas:

- *Low Carbon:* For example, large-scale renewable generation, distributed energy resources (DER), electric vehicles (EV), and carbon capture and sequestration (CCS).
- *Grid Performance:* For example, advanced distribution and substation automation



(Source: General Electric)

Distribution Systems, Substations, and Integration of Distributed Generation. Figure 36
Smart Grid Technologies span across the entire electric grid (Source: General Electric)

(self-healing); wide-area adaptive protection schemes (special protection schemes); wide-area monitoring and control systems (power management unit [PMU]-based situational awareness); asset performance optimization and conditioning (condition based monitoring); dynamic rating; advanced power electronics (e.g., flexible AC transmission system (FACTS), intelligent inverters, etc.), high temperature superconducting (HTS), and many others.

- *Grid-Enhanced Applications:* For example, distribution management systems (DMS); energy management systems (EMS); outage management systems (OMS); demand response (DR); advanced applications to enable active voltage and reactive power management (integrated voltage/VAR control (IVVC), coordinated voltage/VAR control (CVVC)); advanced analytics to support operational, non-operational and BI decision making; distributed energy resource management; microgrid and virtual power plant (VPP); work force management; geospatial asset management (geographic information system (GIS)); key performance indicator (KPI) dashboards and advanced visualization; and many others.
- *Customer:* For example, advanced metering infrastructure (AMI); home/building automation (home automation network (HAN)); energy management systems and display portals; electric vehicle (EV) charging stations; smart appliances, and many others.
- *Cyber Security and Data Privacy*
- *Communication and Integration Infrastructure*

Distributed Generation Technology Landscape

Common types of distributed generation include:

- Non-renewable generation:
  - Combustion turbine generators
  - Micro-turbines
  - Internal combustion
  - Small steam turbine units
- Renewable generation:
  - Low/high temperature fuel cells (e.g., alkaline fuel cell (AFC), molten carbonate fuel cell

(MCFC), phosphoric acid fuel cell (PAFC), polymer electrolyte membrane fuel cell (PEMFC), solid oxide fuel cell (SOFC), direct methanol fuel cell (DMFC))
- Photovoltaic (PV) (mono-, multicrystalline)
- Concentrated PV (CPV)
- Thin-film solar
- Solar thermal
- Hydro-electric (e.g., run-of-river)
- Wind/mini-wind turbines
- Tidal/wave
- Ocean thermal energy conversion (OTEC) (e.g., land-, shelf-, floating-based plants, open, close, and hybrid cycles)
- Energy storage (in the dispatch mode operations)

Each of these technologies is characterized by different electric efficiency, performance, installation footprint, and capital and operational costs.

Demand-Response Design and Operational Challenges

Demand-response (DR) interconnection engineering and engineering details depend on the specific installation size (kW vs. MW); however, the overall components of the installation should include the following:

- DG prime mover (or prime energy source) and its power converter
- Interface/step-up transformer
- Grounding (when needed – grounding type depends on utility-specific system requirements)
- Microprocessor protective relays for:
  - Three-, single-phase fault detection and DG overload (50, 51, 51V, 51N, 59N, 27N, 67)
  - Islanding and abnormal system conditions detection (81o/u, 81R, 27, 59)
  - Voltage and current unbalances detection (46, 47)
  - Undesirable reverse power detection (32)
  - Machine-based DG synchronization (25)
- Disconnect switches and/or switchgear(s)
- Metering, control, and data logging equipment
- Communication link(s) for transfer trip and dispatch control functions (when needed)

Table 3 summarizes the common DG interconnection requirements of utilities for various DG sizes (some details will vary based on utility-specific design and engineering practices).

Demand-Response Integration and "Penetration" Level

Integration of DG may have an impact on system performance. This impact can be assessed based on:

Distribution Systems, Substations, and Integration of Distributed Generation. Table 3 DG interconnection requirements of utilities

Requirements	DG less than 10 kW	DG 10–100 kW	DG 100–1,000 kW	DG > 1,000 kW or >20% feeder load
Disconnect switch	Yes	Yes	Yes	Yes
Protective relays: islanding prevention and synchronization	Yes	Yes	Yes	Yes
Other protective relays (e.g., unbalance)	Optional	Optional	Yes	Yes
Dedicated transformer	Optional	Optional	Yes	Yes
Grounding impedance (due to ground fault contribution current)	No	No	Optional	Often
Special monitoring and control requirements	No	Optional	Yes	Yes
Telecommunication and transfer trip	No	Optional	Optional	Yes

- Size and type of DG design: power converter type, unit rating, unit impedance, relay protection functions, interface transformer, grounding, etc.
- Type of DG prime mover: wind, PV, ICE (internal combustion engine), current transformer, etc.
- Interaction with other DG(s) or load(s)
- Location in the system and the characteristics of the grid, such as:
  - Network, auto-looped, radial, etc.
  - System impedance at connection point
  - Voltage control equipment types, locations, and settings
  - Grounding design
  - Protection equipment types, locations, and settings
  - Others

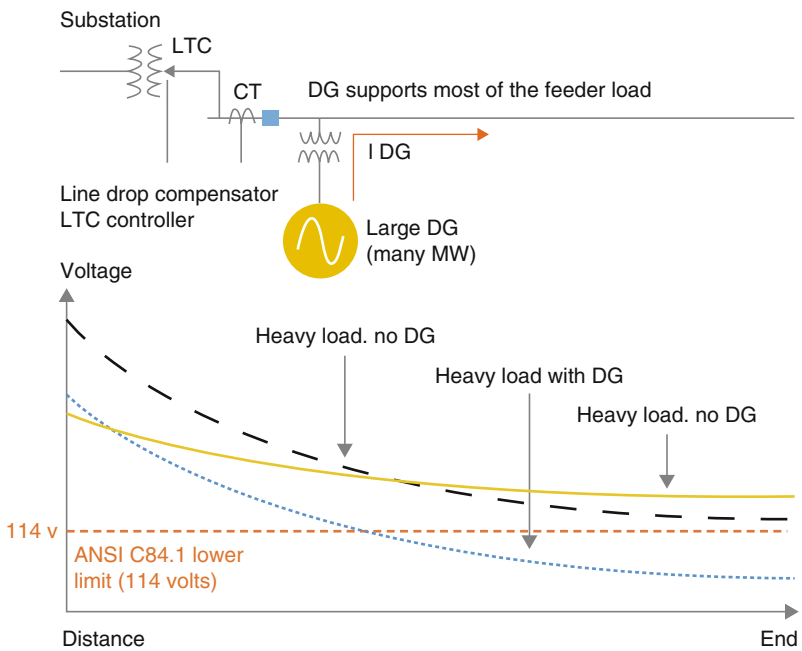
DR system impact is also dependent on the "penetration" level of the DG connected to the grid. There are a number of factors that should be considered when evaluating the penetration level of DG in the system. Examples of DG penetration level factors include:

- DG as a percent of feeder or local interconnection point peak load (varies with location on the feeder)
- DG as a percent of substation peak load or substation capacity
- DG source fault current contribution as a percent of the utility source fault current (at various locations)

Distributed Generation Impact on Voltage Regulation

Voltage regulation, and in particular voltage rise effect, is a key factor that limits the amount (penetration level) of DG that can be connected to the system. Figure 37 shows an example of the network with a relatively large (MW size) DG interconnected at close proximity to the utility substation.

Careful investigation of the voltage profile indicates that during heavy-load conditions, with connected DG, voltage levels may drop below acceptable or permissible by standards. The reason for this condition is that relatively large DG reduces the circuit current value seen by the load tap changer (LTC) in the substation (DG current contribution). Since the LTC sees "less" current (representing a light load) than the actual



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 37
DG connection close to the utility substation

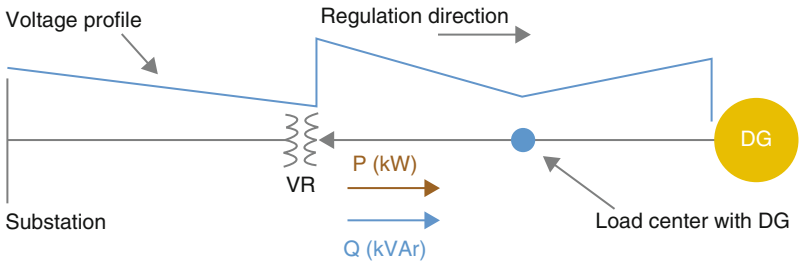
value, it will lower the tap setting to avoid a "light-load, high-voltage" condition. This action makes the actual "heavy-load, low-voltage" condition worse. As a general rule, if the DG contributes less than 20% of the load current, then the DG current contribution effect will be minor and can probably be ignored in most cases.

Figures 38 and 39 show examples of the network with DG connected downstream from the bidirectional line voltage regulator (VR). During "normal" power flow conditions (Fig. 38), the VR detects the real power (P) flow condition from the source (substation) toward the end of the circuit. The VR will operate in "forward" mode (secondary control). This operation is as planned, even though the "load center" has shifted toward the voltage regulator.

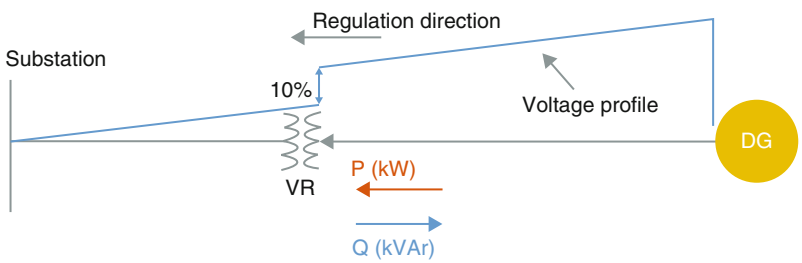
However, if the real power (P) flow direction reverses toward the substation (Fig. 39), the VR will operate in the reverse mode (primary control). Since the voltage at the substation is a stronger source than the voltage at the DG (cannot be lowered by VR), the VR will increase the number of taps on the secondary side. Therefore, voltage on the secondary side increases dramatically.

Distributed Generation Impact on Power Quality

Two aspects of power quality are usually considered to be important during evaluation of DG impact on system performance: (1) voltage flicker conditions and (2) harmonic distortion of the voltage. Depending on the particular circumstance, a DG can either decrease or increase the quality of the voltage received by other users of the distribution/medium-voltage network. Power quality is an increasingly important issue and generation is generally subject to the same regulations as loads. The effect of increasing the grid fault current by adding generation often leads to improved power quality; however, it may also have a negative impact on other aspects of system performance (e.g., protection coordination). A notable exception is that a single large DG, or aggregate of small DG connected to a "weak" grid may lead to power quality problems during starting and stopping conditions or output



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 38
VR bidirectional mode (normal flow)



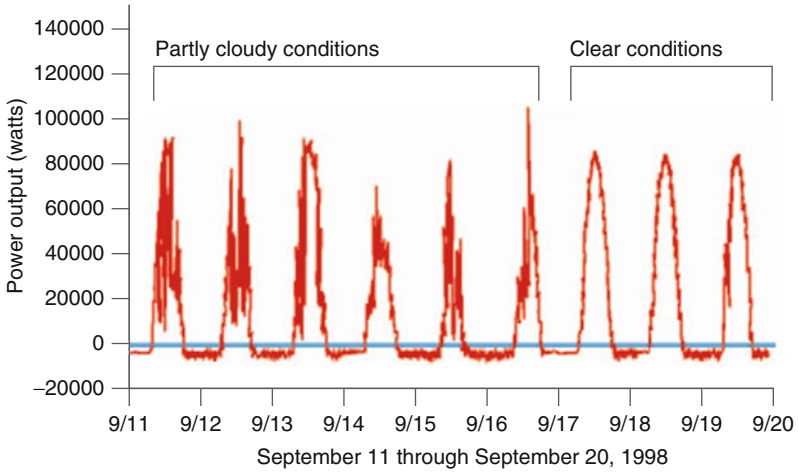
Distribution Systems, Substations, and Integration of Distributed Generation. Figure 39
VR bidirectional mode (reverse flow)

fluctuations (both normal and abnormal). For certain types of DG, such as wind turbines or PV, current fluctuations are a routine part of operation due to varying wind or sunlight conditions (Fig. 40).

Harmonics may cause interference with operation of some equipment, including overheating or de-rating of transformers, cables, and motors, leading to shorter life. In addition, they may interfere with some communication systems located in close proximity of the grid. In extreme cases they can cause resonant overvoltages, "blown" fuses, failed equipment, etc. DG technologies must comply with pre-specified by standards harmonic levels (Table 4).

In order to mitigate harmonic impact in the system, the following can be implemented:

- Use an interface transformer with a delta winding or ungrounded winding to minimize injection of triplen harmonics.
- Use a grounding reactor in neutral to minimize triplen harmonic injection.



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 40
Power output fluctuation for 100 kW PV plant

Distribution Systems, Substations, and Integration of Distributed Generation. Table 4 IEEE 519–1992, current distortion limits for general distribution systems

Maximum harmonic current distortion in % of I_L						
Individual harmonic order (odd harmonics)						
I_{SC}/I_L	<11	11 ≤ h < 17	17 ≤ h < 23	23 ≤ h < 35	35 ≤ h	TDD
<20[a]	4.0	2.0	1.5	0.6	0.3	5.0
20 < 50	7.0	3.5	2.5	1.0	0.5	8.0
50 < 100	10.0	4.5	4.0	1.5	0.7	12.0
100 < 1,000	12.0	5.5	5.0	2.0	1.0	15.0
>1,000	15.0	7.0	6.0	2.5	1.4	20.0

Even harmonics are limited to 25% of the odd harmonic limits. TDD refers to total demand distortion and is based on the average maximum demand current at the fundamental frequency, taken at the PCC

I_{SC} Maximum short circuit current at the PCC, I_L Maximum demand load current (fundamental) at the PCC, h Harmonic number

[a]All power generation equipment is limited to these values of current distortion regardless of $I_{SC}I_L$

- Specify rotating generator with 2/3 winding pitch design.
- Apply filters or use phase canceling transformers.
- For inverters: Specify pulse width modulation (PWM) inverters with high switching frequency. Avoid line-commutated inverters or low switching frequency PWM; otherwise, more filters may be needed.
- Place DG at locations with high ratios of utility short-circuit current to DG rating.

Distributed Generation Impact on Ferroresonance

Classic ferroresonance conditions can happen with or without interconnected DG (e.g., resonance between transformer magnetization reactance and underground cable capacitance on an open phase). However, by adding DG to the system, the case for overvoltage and resonance can increase for conditions such as: DG connected rated power is higher than the rated power of the connected load, presence of large capacitor banks (30–400% of unit rating), during DG formation on a non-grounded island.

DG Impact on System Protection

Some DG will contribute current to a circuit current on the feeder. The current contribution will raise fault levels and in some cases may change fault current flow direction. The impact of DG fault current contributions on system protection coordination must be considered. The amount of current contribution, its duration, and whether or not there are protection coordination issues depend on:

- Size and location of DG on the feeder
- Type of DG (inverter, synchronous machine, induction machine) and its impedance
- DG protection equipment settings (how fast it trips)
- Impedance, protection, and configuration of feeder
- Type of DG grounding and interface transformer

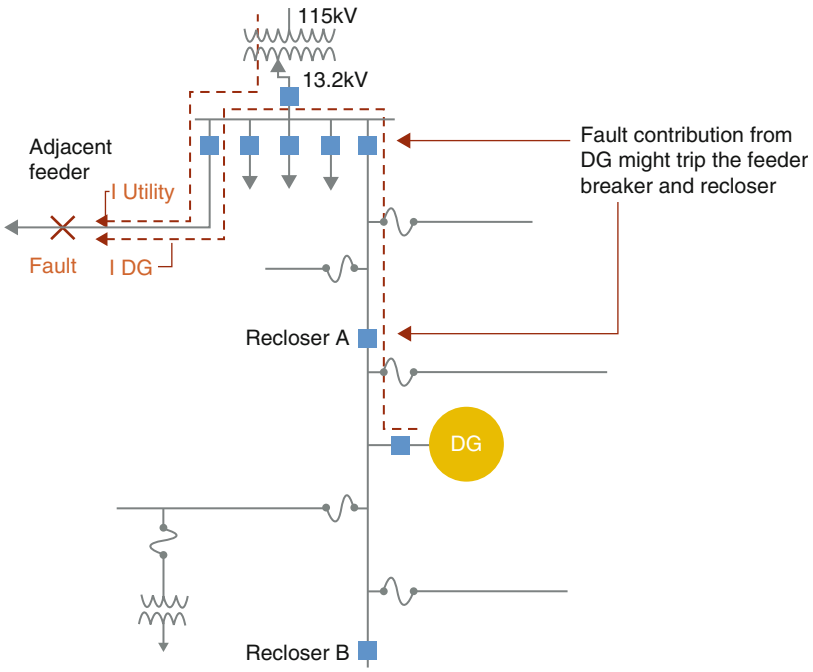
Machine-based DG (IEC, CT, some micro-turbines, and wind turbines) injects fault current levels of four to ten times their rated current with time contribution between 1/3 cycle and several cycles depending on the machine. Inverters contribute about one to two times their rated current to faults and can trip-off very quickly – many in less than one cycle under ideal conditions. Generally, if fault current levels are changed less than 5% by the DG, then it is unlikely that fault current contribution will have an impact on the existing system or equipment operation. Utilities must also consider interrupting capability of the equipment (e.g., circuit breakers, reclosers, and fuses must have sufficient capacity to interrupt the combined DG and utility source fault levels). Examples of DG fault contribution on system operation and possible protection mis-coordination are shown in Figs. 41 and 42.

Future Directions

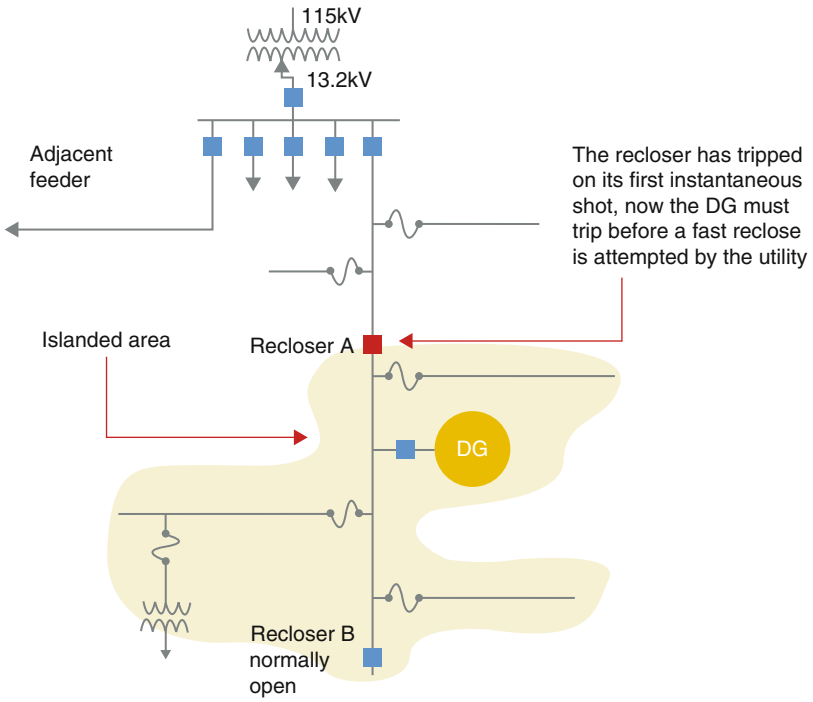
Today's distribution systems are becoming more and more complicated. New methods of producing and storing electrical energy such as PV, fuel cells, and battery storage systems and new methods of consuming electric energy such as smart appliances and plug-in electric vehicles are being connected to the distribution grid. The rate of adoption of these devices will be driven faster as the economic and environmental benefits improve. In response, the automation systems used to monitor, control, and protect them will need to become more sophisticated.

Many utilities are facing these challenges today in increasingly larger areas of their distribution grid. Many developers are building new green communities that contain enough generation and storage to carry the community's load during an outage. Commonly referred to as microgrid operation, these areas of the distribution grid can be momentarily operated while isolated from the rest of main distribution grid. These microgrids can maintain service to un-faulted sections of the grid during some distribution outages. When these microgrids are connected to the distribution system, utilities are also investigating techniques to maximize the value of the distributed energy resources (DER) during times of economic peak or during capacity peaks. The electric energy in these devices can contribute watts, watt-hours, VARs, and voltage or frequency support during times of distribution system need.

The increasing complexity of the distribution grid will force a need to further integrate the various systems on that grid. The various systems described here will become increasingly integrated. These include the FDIR and Volt/VAR systems. As the FDIR system reconfigures



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 41
Undesirable protection trip (back-feeding)



Distribution Systems, Substations, and Integration of Distributed Generation. Figure 42
Unintentional DG Islanding

the distribution system, the Volt/VAR system can then optimize the newly configured feeders. Information such as voltage and VARs from the consumers can be used to improve the amount of the Volt/VAR system can control the grid without violating limits at the consumer. Microgrid operation will further push the integration of all of the distribution systems to maintain a safe, reliable, and efficient distribution grid.

This will have the effect of reducing the costs and increasing the overall benefits of these technologies, while maintaining an improved quality and reliability of the electric energy provided to the customers on the distribution system. This stronger economic justification will drive the rate of advance of these new technologies causing a significant impact on the issues and elements of the design of the distribution system and associated automation systems.

There are vast developments happening in the power industry changing whole transmission and distribution world including substations. Smart grid technologies make their way into transmission and distribution world to improve power supply, make it more efficient and reliable, and decrease greenhouse emissions. This became possible due to rapid developments in power electronics and communications. Major areas of developments are as follows:

- Smart metering implies getting metering data from all possible measurement points including end of the feeder over communications in order to maintain proper distribution voltage level for all consumers. This includes deploying voltage regulators, capacitor banks, switches, and other devices in the distribution network.
- Advances in communications imply real-time, live communication between the consumer, the network, and the generation station, so the utility can balance load demand in both directions. Advanced communications are also needed for manual or automatic reconfiguration of the network in case some components of the network are experiencing failures or deficiencies.
- Advanced protective relays and other control devices with enhanced communications and algorithms capabilities to automatically detect, isolate, and reconfigure the grid to maintain uninterruptable power supply.
- Renewable energy sources will continue fast deployment in the distribution network. Wind power, solar power, hydro power units will increase their capacity and output; energy storage systems will be deployed to help system to meet peak demand and offload system generators. Microgrids will provide consumers with reliable and high-quality energy when connection with a transmission n system is lost or in case of isolated community.

A growing number of electric utilities worldwide are seeking ways to provide excellent energy services while becoming more customer focused, competitive, efficient, innovative, and environmentally responsible. Distributed generation is becoming an important element of the electric utility's smart grid portfolio in the twenty-first century. Present barriers to widespread implementation are being reduced as technologies mature and financial incentives (including government and investor supported funding) materialize. However, there are still technical challenges that need to be addressed and effectively overcome by utilities. Distributed generation should become part of a utility's day-to-day planning, design, and operational processes and practices, with special consideration given to:

- Transmission and distribution substation designs that are able to handle significant penetration of distributed generation
- Equipment rating margins for fault-level growth (due to added distributed generation)
- Protective relays and settings that can provide reliable and secure operation of the system with interconnected distributed generation (can handle multiple sources, reverse flow, variable fault levels, etc.)
- Feeder voltage regulation and voltage-drop design approaches that factor possible significant penetration of distributed generation
- Service restoration practices that reduce the chance of interference of distributed generation in the process and even take advantage of distributed generation to enhance reliability where possible
- Grounding practices and means to control distributed generation-induced ground-fault overvoltages.

Bibliography

Primary Literature

1. Smart Grids, European Union, http://www.smartgrids.eu/
2. United States Department of Energy, Office of Electric Transmission and Distribution (2003) Grid 2030: a national vision for electricity's second 100 years
3. EPRI (2009) Report to NIST on the smart grid's interoperability standards roadmap

Books and Reviews

Institute of Electrical and Electronics Engineers (1994) IEEE recommended practice for electric power distribution for industrial plants, IEEE Std 141-1993. Institute of Electrical and Electronics Engineers, New York. ISBN 978-1559373333

Killian C, OG&E, Flynn B (2009) Justifying distribution automation at OG&E. In: 2009 DistribuTECH conference, San Diego, 3–5 Feb 2009

Lathrop S, Flynn B, PacifiCorp (2006) Distribution automation pilot at PacifiCorp. In: Western Energy Institute, 2006 operations conference, Costa Mesa, 5–7 Apr 2006

Lemke JW, Duke Energy, Flynn B (2009) Distributed generation, storage and the smart grid. In: Autovation 2009, The utilimetrics smart metering conference and exposition, Denver Colorado, 13–16 Sep 2009

Stewart R, Flynn B, Pepco Holdings Incorporated (2007) Modeling DA improvements to reliability performance metrics. In: 2007 WPDAC, Spokane, Washington, DC, 3–5 Apr 2007

Weaver T, AEP, Flynn B (2011) Achieving energy efficiency through distribution system improvements: integrated volt/VAR control (IVVC). In: 2011 DistribuTECH conference, San Diego, 3–5 Feb 2011

Westinghouse Electric Corporation (1959) Electric utility engineering reference book, Distribution systems, vol 3. Westinghouse Electric Corporation, Pittsburgh

Dosimetry

John W. Poston, Sr.
Department of Nuclear Engineering, Texas A & M University, College Station, TX, USA

Article Outline

Glossary
Definition
Introduction
Thermoluminescence Dosimetry
Estimates of Absorbed Doses to Organs of the Body
Thermoluminescence Dosimetry for Neutron Radiations
Optically Stimulated Luminescence
Electronic Dosimeters
Summary
Future Directions
Bibliography

Glossary

Absorbed dose The amount of energy deposited by ionizing radiation per unit mass of the material. Usually expressed in the special radiologic unit rad or in the SI unit the gray (Gy). One Gy equals 1 J/kg or 100 rad.

Dosimeter Any device worn or carried by an individual to establish total exposure, absorbed dose, or equivalent (or the rates) in the area or to the individual worker while occupying the area.

Equivalent dose (Formally the dose equivalent) The product of the absorbed dose and the radiation-weighting factor (formerly the quality factor) for the type of radiation for which the absorbed dose is measured or calculated. The equivalent dose is used to express the effects of radiation-absorbed dose from many types of ionizing radiation on a common scale. The special radiologic unit is the rem or in the SI unit the sievert (Sv). One sievert is equal to 1 J/kg or 100 rem.

Exposure A quantity defined as the charge produced in air by photons interacting in a volume of air of known mass. An old quantity that is generally no longer used. Also, a general term used to indicate any situation in which an individual is being irradiated.

Ionization The process of removing one or more electrons from an atom or a molecule. The positively charged atom and the negatively charged electron are called an *ion pair*.

Isotope One of two or more atoms with the same number of protons but a different number of neutrons in their nuclei. A radioisotope is an isotope of a chemical element that is unstable and transforms by emission of nuclear particles and electromagnetic radiation to reach a more stable state. This term is often misused because unless the materials

are the same element this term should not be used (see radionuclide below).

Nuclide A general term to indicate an atomic nucleus characterized by its atomic number (number of protons), number of neutrons, atomic mass, and energy state.

Radiation Used in this section to mean ionizing radiation. That is, particles or electromagnetic radiation emitted from the nucleus with sufficient energy to cause ionization of atoms and molecules composing the material with which the radiation is interacting.

Radionuclide A nuclide that is radioactive and, upon decaying, emits ionizing radiation.

Definition

Dosimetry is best defined as "the theory and application of principles and techniques associated with the measurement of ionizing radiation" [1].

Introduction

The term "dosimetry" can be best explained by assuming it was derived from combining two words: "dose" and "measurement." The word dose is shorthand for several quantities associated with the profession of health physics (i.e., radiation protection and safety). The terms include the "absorbed dose," which is a measure of the energy deposited per unit mass of material, and the "equivalent dose," which includes consideration of the biological effects of different radiations, when the same absorbed dose is delivered to matter. The term "equivalent dose" is now used instead of the older term dose equivalent to signify changes in the ICRP recommended radiation and tissue weighting factors. There are many other "dose terms" used in health physics but these will not be included here because the fundamental quantity associated with dosimetry is the absorbed dose. Of course, the term measurement implies the use of some sort of detector that is sensitive to the ionizing radiation being measured. These "detectors" can take many forms from photographic film, first used more than 100 years ago, to sophisticated solid-state detectors being introduced today.

Scientists have been detecting radiation for more than a century using a wide variety of detectors. Initially, the detectors were either photographic film or simple ionization chambers filled with air. Crude scintillation systems led to the invention of detectors such as the Geiger–Mueller counter and more sophisticated proportional counters and detectors designed for specific applications and/or to detect a specific radiation. A discussion of these detectors would fill a textbook [2–5] and see also entry C. Radiation Detection Devices (in this encyclopedia). For this reason, this discussion of dosimetry will focus on two of the more modern dosimeters used to monitor the absorbed dose to occupationally exposed workers in nuclear facilities across the United States.

As indicated above, *Dosimetry* is best defined as "the theory and application of principles and techniques associated with measurement of ionizing radiation" [1]. In reality, two basic areas encompass the term "dosimetry." These are called external dosimetry and internal dosimetry. Again, these terms are shorthand descriptions of the more complex exposure conditions being considered. External dosimetry simply means the measurement of radiation that exists outside the human body. Basically, this type of dosimetry uses radiation detection devices and instrumentation to establish the characteristics of the radiation field. These measurements provide information in many forms, for example, the energy or energy spectrum of the radiation, the radiation intensity, the types of radiation present, and other useful information. In many cases, the radiation detectors used for these measurements are called "dosimeters"; an indication that the sole purpose is to measure the radiation-absorbed dose and which leads to an estimate of the equivalent dose. It is important to remember that, because the radiation source and the dosimeters are both outside the body, the measurement does not provide a direct measurement of the absorbed dose to the organs of the body. Methods used to provide estimates of the absorbed doses to organs and tissues of the body will be discussed later.

When a radionuclide or radionuclides are taken into the body, through inhalation, ingestion, injection, or assimilation through the intact skin, there is a completely different set of challenges facing the dosimetrist. Internal dosimetry is defined as "a process of measurement and calculation that results in an estimate of the absorbed dose to organs and tissues of the

body from an intake of radioactive material" [1]. Internal dosimetry is primarily confined to the use of mathematical models and calculational techniques based on an internationally agreed upon set of standard assumptions. The dose estimate relies on mathematical models that describe the uptake, distribution, and retention of the radioactive material in the body. However, the calculations may be based on a set of measurements, such as the concentration of airborne radioactivity in a work area, the activity of radioactive material deposited in the body or specific organs in the body, or measurement of the concentration of radioactivity in excreta, such as urine or feces. Even with these measurements as initial input, internal dosimetry must rely on models of a reference human and calculational techniques. These aspects will not be discussed here.

This section will focus on a discussion of external dosimetry methods, which are primarily used to monitor radiation exposures of occupationally exposed workers conducting licensed activities in the US. Radiation dosimeters that are no longer widely used, such as film badges and pocket ionization chambers will not be discussed.

Thermoluminescence Dosimetry

In 1950, Daniels suggested that the thermoluminescence (TL) phenomenon could be used as a radiation dosimeter [3]. This suggestion came late in the development of radiation dosimeters even though it was known that Henri Becquerel, as well as his father, had mentioned this phenomenon in his scientific papers. In addition, the relation between X-ray exposure and thermoluminescence was observed as early as 1904. Nevertheless, after many struggles and failures in the research of Cameron and his colleagues, thermoluminescence and thermoluminescence dosimetry (TLD) became a reality and flourished in the late 1960s and 1970s [3]. For a very long time, TLD has been the most popular method of personnel monitoring.

In these dosimeters, the absorbed dose is determined by observing the emitted light from an inorganic crystal after exposure to radiation. The light is released from the crystal as it is heated under controlled conditions. The heat energy originally was provided by electrical heating but subsequent developments in TLD led to the use of high-intensity light as an alternate method. Regardless of the method of heating, the amount of light emitted is directly proportional to the radiation energy deposited in the TL material. This light is normally measured with a photomultiplier tube sensitive to the wavelength of the emitted light. It must be remembered that the TLDs are not "absolute dosimeters" and, therefore, require proper calibration in the radiation fields to which the dosimeters will be exposed.

Detailed explanations of the TL phenomenon have been offered by a number of scientists but a simple bandgap model can be used to explain the basic mechanism. The usual procedure is to refer to the energy-level diagram in an insulating crystal. In a pure crystal, radiation impinging on the crystal would free electrons and these electrons would pass from the valence band to the conduction band. These electrons would not remain in the conduction band for a long period and would return to the valence band releasing the energy acquired in the form of light. In a pure crystal, this light would be absorbed and would not escape the crystal. In TLDs, dopants (impurities) are added to the crystal and these impurities reside in the forbidden or bandgap between the valence and conduction bands. When these crystals are exposed to radiation, the loss of electrons from the valence band creates positively charged atoms ("holes"). The electrons and holes may migrate through the crystal until they recombine or are "trapped" by the impurity atoms (dopants) residing in the bandgap. Thus, the energy absorbed by the crystal is stored until it is released, in the form of light, through heating the crystal (thus, thermoluminescence). This light, which is now characteristic of the impurity sites, can escape the crystal and can be measured with an external detector (i.e., a photomultiplier tube).

It is important to realize that these trapping sites may exist at many different levels in the bandgap and it is not correct to assume that all electrons (or holes) are trapped at exactly the same energy level. Thus, the light intensity may vary as a function of temperature and the plot of the light intensity as a function of temperature (called a "glow curve") may exhibit a number of peaks and valleys depending on the number of trapping levels in the crystal. Either the total light emitted or the height of a particular peak may be used to determine the absorbed dose (upon proper calibration). It is also important that the heating cycle be very reproducible to avoid causing fluctuations in the peak heights.

There are a large number of inorganic materials that have been studied for use as TLDs. Table 1 presents a summary of the characteristics of some of the most popular materials (but there are many other possible TLD materials). In dosimetry, it is common to use materials that are "tissue equivalent" in terms of the interactions of photons or other radiations with the dosimeters. Thus, the closer the effective atomic number of the material is to that of tissue (~7.6), the more tissue equivalent is the material. For historical reasons, LiF is the standard to which all other TLD materials are compared. The standard LiF is the natural form of lithium with the normal concentrations of the isotopes of Li-6 (7.4%) and Li-7 (92.6%). Also in this table are listed the temperatures of the "main peak." This designation is the peak in the TLD glow curve that is used to determine the absorbed dose. One big disadvantage of TLDs is that the dosimeter can only be read (evaluated) once. Heating the crystal essentially releases all the electrons or holes that are trapped and an opportunity to confirm the reading is not possible.

Dosimetry. Table 1 Summary of characteristics of thermoluminescence dosimetry (TLD) materials

TLD material	Effective atomic number (Z_{eff})	Temperature of main peak
$CaSO_4$:Mn	15.3	110°C
$CaSO_4$:Dy	15.5	220°C
CaF_2:Mn	16.3	260°C
CaF_2:Dy	16.3	180°C
LiF:Mg,Ti[a]	8.2	195°C
$Li_2B_4O_7$:Mn	7.4	200°C
Al_2O_3:C	10.2	185°C

[a]LiF:Mg,Ti is the standard material to which all other TLD materials are compared

Table 2 compares other characteristics of the TLD materials. These include the "light output" of the material when exposed to ^{60}Co radiation compared to the light output for the standard, that is, LiF. The data for $Li_2B_4O_7$:Mn are somewhat misleading because the measurements quoted in this table were made with the standard photomultiplier tube used for all other TLD materials. However, because the wavelength of light from the $Li_2B_4O_7$:Mn is different, the output can be improved significantly by replacing the normal photomultiplier with one with a photocathode sensitive to the correct wavelength of light. The energy response is the ratio of the light output at energy of 30 keV to that from irradiation with ^{60}Co. Except for $Li_2B_4O_7$:Mn, most materials overrespond to low-energy photon radiation.

As can be seen in Table 2, the usual TLD materials are very sensitive to radiation with lower limits of detection in the range of tenths of microgray. Upper limits range from only 10 Gy to more than 10^4 Gy. The

Dosimetry. Table 2 Summary of dosimetric characteristics of TLD materials

TLD material	Efficiency to Co-60	Energy response	Useful dose range	Fading
$CaSO_4$:Mn	70	~10	0.2 μGy–10^2 Gy	50% in 24 h
$CaSO_4$:Dy	20	~12.5	0.2 μGy–10^3 Gy	2% in 1 month 8% in 6 months
CaF_2:Mn	10	~13	10 μGy–3 × 10^3 Gy	10% in 16 h 15% in 2 weeks
CaF_2:Dy	30	~12.5	0.1 μGy–10^4 Gy	10% in 24 h 16% in 2 weeks
LiF:Mg,Ti	1.0	1.25	10 μGy–3 × 10^3 Gy	5% in 1 year
$Li_2B_4O_7$:Mn	0.15	0.9	0.5 μGy–10^4 + Gy	<5% in 3 months
Al_2O_3:C	70	2.9	0.5 μGy–10 Gy	<3% in 1 year

term "fading" is an indication of the ability of the TLD material to retain the stored energy and thus the stored information necessary to assign the absorbed dose from the wearing of the dosimeter. As is shown, LiF:Mg,Ti, $Li_2B_4O_7$:Mn, and Al_2O_3:C have good energy storage capability, which has led to a focus on these three materials.

Estimates of Absorbed Doses to Organs of the Body

In the United States, federal regulations require the reporting of three quantities for all occupationally exposed workers who are anticipated to receive doses in excess of 10% of the federal limits. These quantities are the "deep-dose equivalent," the "eye-dose equivalent," and the "shallow-dose equivalent." The deep-dose equivalent is defined as the dose at 1-cm depth in the body, which produces an overestimate of the absorbed doses because most organs and tissues of the body are located deeper than 1 cm (or 1,000 mg/cm^2). The eye-dose equivalent considers the dose to the lens of the eye, which is assumed to be at a depth of 300 mg/cm^2. Finally, the shallow-dose equivalent (or more properly the skin-dose equivalent) is assumed to be at a depth of 7 mg/cm^2.

Now the question arises, "How does one measure these absorbed doses, and the subsequent equivalent doses, with radiation dosimeters located outside the body of the worker?" The approach taken has been used for many years and is not new. It has been applied since the Manhattan Project era and is an accepted method to provide these dose estimates. The technique involves using multiple detector elements, that is, typically four TLDs, and covering these TLDs with different thicknesses of materials (called filters) to represent these depths. So, the TLD designated to measure the deep-dose equivalent is covered with a material having a density thickness of 1,000 mg/cm^2. The eye-dose equivalent is determined by covering the TLD with material with a density thickness of 300 mg/cm^2. Usually, there are two different materials of this density thickness in the dosimeter. Finally, there is a thin filter included to allow extrapolation to the depth of 7 mg/cm^2. It is very difficult to provide a direct measurement at such an extremely shallow depth.

Thermoluminescence Dosimetry for Neutron Radiations

TLDs have their primary application in dosimetry for X-ray and gamma-ray fields. In addition, the TLDs have limited sensitivity to beta radiation. Because certain materials in the TLDs interact with neutrons, TLDs can be used to measure both thermal and fast neutron dose – with proper calibrations. Table 3 lists the pertinent information regarding three types of LiF TLDs as these are applied to neutron dosimetry. As can be seen in this table, the natural LiF TLD has the normal concentrations on Li-6 and Li-7. This material is designated as TLD-100. The other two materials are designated TLD-600 and TLD-700. TLD-600 contains a high concentration of the isotope Li-6 with less that 5% Li-7. TLD-700 contains essentially all the isotope Li-7 with a very small amount of Li-6. Notice also the differences in the thermal neutron cross sections (probability to absorb neutrons) for these two isotopes. These differences play a role in the dosimetry of both thermal and fast neutrons.

Thermal neutron dosimetry is based on the "difference technique" used to separate the photon and thermal neutron dose from each other. This technique is similar in some ways to the standard method using bare and cadmium-covered gold foils to measure the thermal neutron fluence in a nuclear reactor core. Basically, the LiF TLD-600 is sensitive to photon radiation as well as to thermal neutron radiation. The LiF TLD-700 has the same photon sensitivity but essentially no sensitivity to thermal neutrons. Thus, when used in a mixed photon and thermal neutron field the TLD-600 will provide the absorbed dose for both the photons and the thermal neutrons. The TLD-700 will provide only the absorbed dose from the photon radiation and the

Dosimetry. Table 3 Characteristics of lithium fluoride TLDs for neutron dosimetry

TLD type	Li-6 percentage	Li-7 percentage	Thermal neutron cross section
Natural (TLD-100)	7.4%	92.6%	N/A
TLD-600	95.62%	4.38%	950 barns
TLD-700	0.007%	99.993%	0.033 barns

difference between the doses indicated by the two dosimeters is the thermal neutron dose.

Fast neutron dosimetry uses a similar technique but to obtain the fast neutron dose the "albedo technique" is used. The albedo technique relies upon the dosimeter being held closely to the body and the fast neutron dose is measured as the fast neutrons enter the body, are moderated there by tissue, are subsequently reflected from the body and hit the dosimeter. There are many designs of albedo fast neutron dosimeters but the concept is the same as that outlined above. The fast neutron dose is obtained by using the difference technique as before.

Note that using TLDs to measure either thermal neutron or fast neutron dose require very careful calibration of the dosimeters in radiation fields approximating those in which the exposures are anticipated. In addition, in very high photon radiation fields with a low percentage of thermal or fast neutrons, these dosimeters may provide data that is highly suspect. This is often the consequence of subtracting two very large numbers (large photon dose) to obtain an estimate of the very low thermal or fast neutron dose. Other methods of neutron dosimetry, for example, track-etch detectors, may be preferred in these situations.

Optically Stimulated Luminescence

Currently, the dosimetry method of choice for dosimetry appears to be optically stimulated luminescence (OSL). Even though film and TLD are still used to some extent, many facilities are switching over to this newer technology. OSL may be used, not only for personnel monitoring, but also for environmental monitoring and medical dosimetry. Basically, OSL is very similar to TLD in terms of the basic physics associated with the energy deposition, storage, and release. The major difference is that, instead of using heat, laser light is used to release ("detrap") the electrons. The laser is pulsed at a rate of 4,000 times per second and is directed to only a small area on the material. This provides an opportunity for multiple readings on the same dosimeter, if necessary, as the laser can be focused on another region of the crystal. In a similar fashion to the development of TLDs in the latter part of the twentieth century, many materials have been studied for possible use as OSL dosimeters. These materials include halides, sulfates, sulfides, and oxides [6].

However, the material of choice is an Al_2O_3:C crystalline detector. Single crystals of Al_2O_3:C are ground into a powder and mixed with a polyester base. This mixture is deposited on a polyester film about 0.03 cm thick, which can be fabricated in a thin strip for incorporation into a dosimeter. This material has a good response to photon radiation as well as a response to beta radiation. Copper (0.18 g/cm^2) and tin (0.39 g/cm^2) filters, as well as an open area, are used in the dosimeter (as described above) to provide the dosimetry quantities of interest. Commercially available dosimeters have a dose measurement range for photons of 1 mrem to 1,000 rem (10 μSv to 10 Sv) over an energy range from 5 keV to more than 40 MeV. For beta radiation, the dose measurement range is from 10 mrem to 1,000 rem (100 μSv to 10 Sv) over an energy range of 150 keV to 10 MeV (average energy). The commercially available OSL dosimeters may be used for up to 1 year. If the packaging is not compromised, the dosimeter is unaffected by heat, moisture, and pressure [7].

Electronic Dosimeters

Currently, in many situations, such as in a nuclear power plant, it is common to wear two types of dosimeters. One of these dosimeters is usually a TLD or an OSL-type. This dosimeter is usually designated as the "dosimeter of record." That is, these dosimeters are worn for long periods of time (i.e., 1 month, 3 months, or perhaps 1 year) and provide a measure of the total dose received by the exposed worker over the wearing period. It is these doses that are reported annually to the US Nuclear Regulatory Commission, as required by the federal regulations. The second dosimeter is a modern, electronic dosimeter that may contain as many as three small detector elements (usually solid-state detectors such as silicon diodes). Electronic dosimeters are used to monitor the work and may be worn for short periods of time. The primary function of these dosimeters is work and dose control. The dosimeters feature adjustable alarm points that may be set by a computer, before use, based on the anticipated total dose received or maximum dose rate encountered. Workers are trained to recognize the alarms and understand the proper response to these alarms.

There are many different electronic dosimeters but most have similar characteristics. A typical dosimeter would use small silicon diode detectors, which would be sensitive to both photons (50 keV to 6 MeV) and beta radiation (>60 keV up to more than 2 MeV). Doses from 0.1 mrem (1 μSv) to 10 Sv can be measured with dose rates ranging from 0.01 mrem/h (0.1 μSv/h) to 1,000 rem/h (10 Sv/h). Most dosimeters feature both audible and visible alarms. Typically, these dosimeters are lightweight, from 50 g up to perhaps 200 g. A unique feature of some of these dosimeters, and a good radiation protection practice, is the use of permanent stations throughout the plant that interrogate the dosimeters as the worker passes by the station and transmits this information to a central station. Other types of dosimeters contain small transmitters that transmit the accumulated dose (or dose rates) to central locations, which are monitored by the radiation safety staff. As technology moves forward, it is difficult to predict what the future holds in terms of the next generation of dosimeters.

Summary

The measurement of radiation energy deposited in material, that is, the measurement of the absorbed dose, is the primary goal of the practice of dosimetry. Over the last 100 years or more, dosimetry has taken many forms as science and technology have made significant progress. Many of the techniques have been relegated to the history books as other more advanced techniques have been introduced. This short discussion of dosimetry was intended to present the basic concepts and to provide two examples of modern dosimeters used to monitor personnel that are occupationally exposed to ionizing radiation, as well as to introduce the use of electronic dosimeters, which are used widely in nuclear utilities.

Future Directions

Approaches to dosimetry have changed rapidly with developments in electronics and computers. The last decade or so has seen the design and manufacture of dosimeters that are small but incorporate computer capabilities. These dosimeters allow the setting of dose and dose rate alarms, remote interrogation of the dosimeters to monitor worker exposure, and many other features. It appears these trends will continue as the demand for "smarter" dosimeters, with many more capabilities, for use in nuclear facilities as well as in emergency response continues to increase.

Bibliography

1. Poston JW Sr (1987) Dosimetry. In: Encyclopedia of physical science and technology, vol 6. Academic, New York
2. Knoll GF (2000) Radiation detection and measurements, 3rd edn. Wiley, New York
3. Eichholz GG, Poston JW (1979) Principles of nuclear radiation detection. Ann Arbor Science Publishers, Ann Arbor
4. Tsoulfanidis N (1995) Measurement & detection of radiation, 2nd edn. Taylor & Francis, Washington, DC
5. Kase KR, Bjarngard BE, Attix FH (eds) (1985) The dosimetry of ionizing radiation, vol I–III. Academic, New York
6. Boetter-Jensen L, McKeever SWS, Wintle AG (2000) Optically stimulated luminescence dosimetry. Elsevier, Maryland Heights
7. R. S. Landauer, Inc. (2010) http://www.osldosimetry.com/luxel/. Accessed 9 May 2010

Dredging Practices and Environmental Considerations

Craig Vogt[1], Gregory Hartman[2]
[1]Craig Vogt Inc, Hacks Neck, VA, USA
[2]Hartman Associates Inc, Waterway Engineering & Sediment Remediation, Port Orchard, WA, USA

Article Outline

Glossary
Definition of Subject
Introduction
Dredging and Dredged Material Management
Dredging
Transportation of Dredged Material
Dredged Material Disposal and Placement Alternatives
Control of Upstream Sources of Sediments and Contaminants
Future Directions: Sustainable Dredging and Dredged Material Management
Bibliography

Glossary

Beneficial use Placement or use of dredged material as resource materials in productive ways, which provide environmental, economic, or social benefits.

Confined disposal facility (CDF) An engineered structure for containment of dredged material consisting of dikes or other structures that enclose a disposal area above any adjacent water surface, isolating the dredged material from adjacent waters during placement. Other terms used for CDFs that appear in the literature include "confined disposal area," "confined disposal site," and "dredged material containment area."

Contaminant A chemical or biological substance in a form that can be incorporated into, onto, or be ingested by or harm aquatic organisms, consumers of aquatic organisms, or users of the aquatic environment.

Contaminated sediment or contaminated dredged material Contaminated sediments or contaminated dredged materials are defined as those that may cause an unacceptable adverse effect on human health or the environment.

Dredged material Material excavated from fresh, estuarine, or ocean waters. The term "dredged material" refers to material which has been dredged from a water body and disposed in a disposal site. The term "sediment" refers to material on the bed of a water body prior to the dredging process.

Dredging Underwater excavation is called dredging. "Dredging" is the term given to removal by digging, gathering, or pulling out materials from the bed to deepen waterways and to create harbors, channels, and berths. Dredging is also conducted for construction purposes, for mining, and for environmental cleanup and enhancement.

Habitat The specific area or environment in which a particular type of plant or animal lives. An organism's habitat provides all of the basic requirements for the maintenance of life. Typical coastal habitats include beaches, marshes, rocky shores, bottom sediments, mudflats, and the water itself.

Open-water disposal Placement of dredged material in rivers, lakes, estuaries, or oceans via pipeline or surface release from hopper dredges or barges, without confinement.

Sediment Material, such as sand, silt, or clay, suspended in or settled on the bottom of a water body. Sediment input to a body of water comes from natural sources, such as erosion of soils and weathering of rock, or as the result of anthropogenic activities, such as forest or agricultural practices, or construction activities. The term "dredged material" refers to material which has been dredged from a water body, while the term "sediment" refers to material in a water body prior to the dredging process.

Suspended solids Organic or inorganic particles that are suspended in water. The term includes sand, silt, and clay particles as well as other solids, such as biological material, suspended in the water column.

Toxicity Level of mortality or other end point demonstrated by a group of organisms that have been affected by the properties of a substance, such as contaminated water, sediment, or dredged material.

Toxic pollutant Pollutants, or combinations of pollutants, including disease-causing agents, that after discharge and upon exposure, ingestion, inhalation, or assimilation into any organism, either directly from the environment or indirectly by ingestion through food chains, will cause death, disease, behavioral abnormalities, cancer, genetic mutations, physiological malfunctions, or physical deformations in such organisms or their offspring.

Turbidity An optical measure of the amount of material suspended in the water. Increasing the turbidity of the water decreases the amount of light that penetrates the water column. Very high levels of turbidity can be harmful to aquatic life.

Definition of Subject

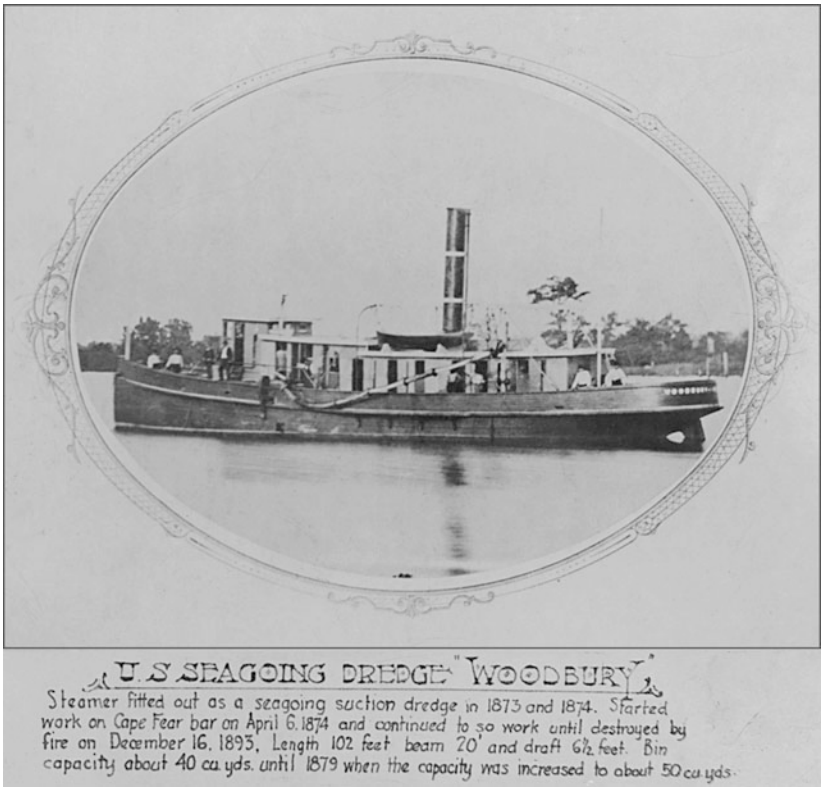
Underwater excavation is called dredging. Dredging is the term given to removal by digging, gathering, or pulling out materials from the bed to deepen waterways and to create harbors, channels, and berths. Dredging is also conducted for construction purposes, for mining, and for environmental cleanup and enhancement. The complete dredging activity includes sediment excavation and removal from the bed, and transport from the dredging site to a disposal area or placement site, which is located in either an open-water, nearshore, or upland location.

Operations that cause potential environmental impacts associated with the dredging process include (1) the sediment removal process from submerged excavation at the point of dredging and (2) the transport and placement for disposal of the dredged material. Environmental concerns relate to the location of the sediment removal by dredging and the disposal or placement site. General environmental considerations include:

- Turbidity in the water
- Acute and chronic toxicity due to chemicals in the dredged material
- Impacts to marine animals
- Loss of habitat
- Air pollutant emissions from the dredge
- Other general construction issues, such as underwater and surface equipment noise

Introduction

Dredges of various designs have been used for many years to create and maintain navigable waterways to move people, goods, and materials. It is theorized that thousands of years ago, blocks of stone that make up the pyramids in Egypt were barged from a distant quarry through a dredged canal. At that time, the canals were likely dredged using a barge with people using long-handled dipper shovels to raise solids out of a waterway and then place those solids on a haul barge deck for disposal elsewhere. Productivity gains likely came about when animal power was used to increase the digging power of early dredges. The late 1800s saw the development of electric and steam power units (Fig. 1), which enabled the construction of huge mechanical dredges with bucket ladders, backhoe dredges, and pipeline dredges with centrifugal pumps.



Dredging Practices and Environmental Considerations. Figure 1
US seagoing dredge "woodbury" – 1873 (Photo courtesy of Corps of Engineers)

Hydraulic technology made great advancements in the 1960s, with the result that hydraulic winches and hydraulic rotary cutter drives became a welcome replacement facilitating the removal of finer-grain sediment (compared to clunky and inefficient mechanical drives) [1].

Today, the dredge type can be hydraulic or mechanical and can be used for a multitude of purposes and projects. The primary purposes are navigation, environmental enhancement, and mining/construction [2].

Navigation Dredging

Most coastal and river ports, harbors, and navigation channels are not naturally deep enough or wide enough to support safe passage of vessels. Navigation channels need to be dredged to create waterway channels with adequate channel area, depth, and access to port and harbor facilities. Nearly all the major ports in the world have at some time required dredging to deepen and widen the access channels, to provide turning basins, and to achieve appropriate water depths to and from waterside facilities.

Virtually, all of the navigation channels created in rivers and harbors have and continue to require maintenance dredging, i.e., the removal of sediments which naturally accumulate on the bottom of the dredged channel. Navigation channel dredging can be categorized as two types. (1) New work dredging is the initial dredging conducted to excavate a channel with navigable depths greater than those that naturally exist. (2) Maintenance dredging is the dredging after new work, which removes accumulated sediments and ensures that the channel continues to provide adequate dimensions for vessels engaged in domestic and international commerce as well as for other types of vessels, such as recreational boating and commercial fishing.

Environmental Enhancement Dredging

In the last 3 decades, dredging has been successfully used to remove contaminated sediments from waterways, with the intention of improving water quality and restoring the health of aquatic ecosystems. Cleanup dredging for removal of contaminants is used in waterways, lakes, ports, and harbors, usually in highly industrialized or urbanized areas that are suffering from past toxic waste and wastewater disposal practices. After removal from the bed, the contaminated sediments are transported and disposed under strict environmental controls (e.g., lined upland confined disposal facilities). In some cases, the contaminated sediments may be treated, and some or all of the sediments used for beneficial objectives. Under proper conditions, a viable alternative to removal is in situ isolation, i.e., the placement of a cap (i.e., a cover of clean material) over the contaminated sediments in their original location.

Dredging is also used for environmental enhancement projects, such as beach nourishment (e.g., replacing lost sand to widen beaches) and providing sediments to enhance marshes and wetlands.

Reclamation, Mining, and Construction Dredging

Dredging is an integral tool in many types of water-related construction projects, such as emplacement of pipelines or immersed tunnels, underwater foundations, and maintaining storage capacity in water supply and recreational reservoirs. In addition, dredging is important in mining activities, with a primary use to provide sand and gravel for construction and reclamation projects. Dredged aggregates have a wide range of uses, including:

- Land reclamation: Pressures arising from population growth have created a need to raise the elevation of low-lying areas for port and infrastructure development and/or to construct new land areas. Such pressures are likely to continue, but loss of marine habitat is an opposing issue when creating new land.
- Construction materials: An increasing quantity of aggregate mined from marine and freshwater borrow sites is used in concrete and fill construction.

Dredging and Dredged Material Management

Dredging and dredged material management consists of the following three elements:

1. Excavation: The dislodgement and removal of sediments (clay, silt, sand, gravel, and rock) from the bed of the water body by a dredge, either mechanically, hydraulically, or by combination of the two dredging methods.
2. Transport: The transport of excavated material from the point of dredging to the final disposal

site. This can be accomplished by haul barges separate from the dredge equipment, or by a dredge equipped with hoppers, or by pipeline from the dredge to the disposal or placement site.

3. Disposal or placement: The final disposal or placement of dredged material. Whether dredged material is disposed or placed (and reused for another purpose, such as creation of a wetland) is determined by a range of factors, including the objectives of the dredging project. The disposal site factors to be considered include sediment type to be dredged (e.g., grain size), location of the dredging project versus the disposal site or beneficial use site, future disposal site utilization, physical and chemical characteristics of the sediment (e.g., is it contaminated?), and available funding.

Increasingly considered a resource, dredged material has a wide number of beneficial use applications that must be considered in dredged material disposal management. This includes consideration for beach nourishment, shoreline fill, and habitat creation or restoration. This concept as a resource is also an important environmental consideration for remedial and cleanup dredging. Dredged material from remedial dredging is usually placed into well-controlled upland confined facilities. Dredged material from clean sediment dredging is often used to cap, cover, and isolate contaminated sediment in the waterway.

Dredging

While specialized dredging equipment varies widely in many sizes and types, dredging is actually accomplished basically by only two dredge types. They are mechanical dredges and hydraulic dredges. The type of dredge is derived from the method of sediment capture and removal from the bed.

Selection of dredging equipment and the methods used to perform the dredging depends on the following factors [3]:

- Physical characteristics of material to be dredged
- Quantities of material to be dredged
- Depth of material to be dredged
- Method of disposal or placement
- Distance to disposal or placement site
- Physical environment of the dredging area(s)
- Physical environment of the disposal area(s)
- Level of contamination of the material to be dredged
- Dredge production capability
- Type of dredges available
- Time, environmental, and economic limits of the project

Mechanical Dredges

Mechanical dredges remove bottom sediment through the direct application of mechanical force to dislodge and excavate the material at almost in situ densities. The mechanical dredges (Fig. 2) are well suited to removing hard-packed material or debris and to working in confined areas, such as in environmental cleanup dredging. Cohesive sediments that are mechanically dredged usually remain intact, with large pieces retaining their in situ density and structure through the dredging and placement process. Sediments excavated with a mechanical dredge are generally placed into a haul barge or scow for transportation from the dredging site to the disposal or placement site.

Mechanical dredges use some form of bucket to excavate and lift the dredged material from the bottom, then load it on to a haul barge or scow. When the haul barge is loaded, a tug or other attendant vessel will take the barge to the disposal or placement site. Mechanical dredges are classified by how the bucket is connected to the dredge. The three standard classifications include:

- Wire rope connected – barge-mounted crane (clamshell or cable bucket)
- Structurally connected – articulated fixed-arm dredge (e.g., backhoe, excavator)
- Chain and structurally connected – bucket ladder

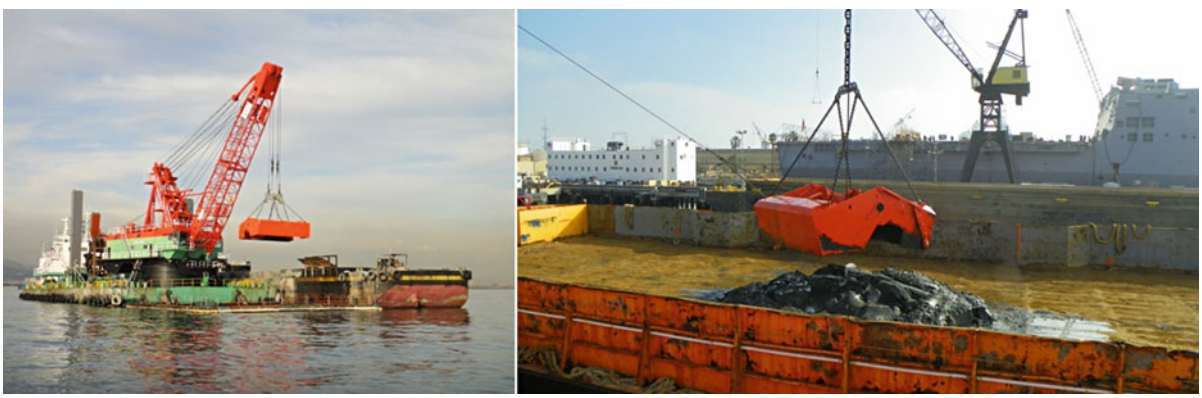
A mechanical dredge is often labeled a clamshell dredge. The clamshell is actually a type of cable-connected bucket used on a mechanical dredge and is the most common type of the mechanical dredges. Mechanical dredges that have the bucket with cable connection to the barge-mounted crane do use a number of different bucket designs for differing sediment characteristics, such as mud, gravel, rock, or boulders.

An articulated fixed-arm mechanical dredge can be a back-acting (backhoe) excavating machine, an advance cut (dipper, excavator) excavating machine, or a bucket (grab) excavating machine.

A bucket ladder dredge is the oldest and the most common type of mechanical dredge. It is primarily used for mining applications. A bucket ladder dredge consists of a large number of buckets linked together in an endless chain which is carried on a ladder that is raised and lowered by hoisting wires. The buckets dig into the face of the cut, and the sediment is carried up the ladder and dumped onto a conveyor belt.

Dredging for environmental cleanup requires much greater precision than navigation dredging and can be accomplished using articulated fixed-arm mechanical dredges, which are similar to conventional upland excavators placed on a barge. The rigid arm, as compared to the cable-connected bucket, provides greater positioning control in placing the bucket on the bottom. Bucket dredges that are designed for a level cut and equipped to be enclosed after the cut are also effective in environmental dredging. These buckets (Fig. 3) minimize the leakage of water and contaminants during the excavation and

Dredging Practices and Environmental Considerations. Figure 2
Mechanical backhoe dredge, New York (Courtesy Great Lakes Dredge & Dock Company)



Dredging Practices and Environmental Considerations. Figure 3
Mechanical dredges: environmental closed buckets (Courtesy Cable Arm Company)

placement of the contaminated material on the barge for transport.

The dragline bucket dredge is similar to the clamshell dredge. They are effective in excavation of gravels, sand, and compact silts. The dragline bucket is not an enclosed bucket and can cause significant turbidity in the water column. Dragline buckets can be operated as a dredge from shore or mounted on a barge. The bucket does not load vertically. Instead, it is lowered to the bottom and then loaded by dragging toward the crane. A dipper dredge is a floating face shovel that digs forward into the face of the excavation and is mounted on a spud barge.

Hydraulic Dredges

Hydraulic dredges are identified by two primary types. They are the pipeline cutterhead dredge and the trailing suction hopper dredge. The hydraulic dredge works by dislodging bed sediment and hydraulic removal of the sediment from the bed of the waterway by suction pipe.

The hydraulic pipeline dredge is generally comprised of the following equipment:

- Cutterhead. The pipeline dredge has an active cutterhead (Fig. 4) that rotates and dislodges the sediment from the bed. This allows the suction, created at the cutterhead by the suction pipe and pump, to be captured and pulled up the suction pipe.
- Suction pipe. The pipe that connects the cutterhead on the bed with the centrifugal pump in the dredge hull.
- Ladder. The ladder raises and lowers the suction pipe and the cutterhead.
- Barge hull. The pipeline dredge is a specialty barge that supports the machinery, pump cable winches, and other equipment in the hull with a lever room, anchor booms, and "walking" spuds on the hull.
- Discharge pipe. The discharge pipe transports the dredged material slurry from the pump to the disposal site.
- Dredge pump(s). The dredge pumps are large centrifugal pumps located on the dredge hull at the waterline and/or on the ladder near the suction mouth.
- Spuds/anchor wires. The spuds and anchor wires are used to anchor the dredge, swing the dredge across the cut, and move the dredge forward.

The pipeline dredge (Fig. 5) is not self-powered. It moves through the cut using the "walking" spud and then the working spud for dredging, thereby allowing

Dredging Practices and Environmental Considerations. Figure 4
Typical hydraulic cutterhead dredge (Courtesy Ellicott Dredges Company)

Dredging Practices and Environmental Considerations. Figure 5
Cutterhead pipeline dredge (CSD), Texas (Courtesy: Great Lakes Dredge & Dock)

the dredge to move forward as it swings the cutterhead from left to right and return.

The pipeline dredge size is classified by the size of the discharge pipe inner diameter. These dredges use hydraulic centrifugal pumps to provide the lifting force to capture and transport the dredged material in a liquid and solid slurry. Pipeline dredges usually work well in loose, "unconsolidated" silts, sands, gravels, and soft clays. In dense consolidated sediments, the hydraulic pipeline dredge depends on an active cutterhead and/or waterjets that can break up the consolidated material at the mouth of the suction pipe.

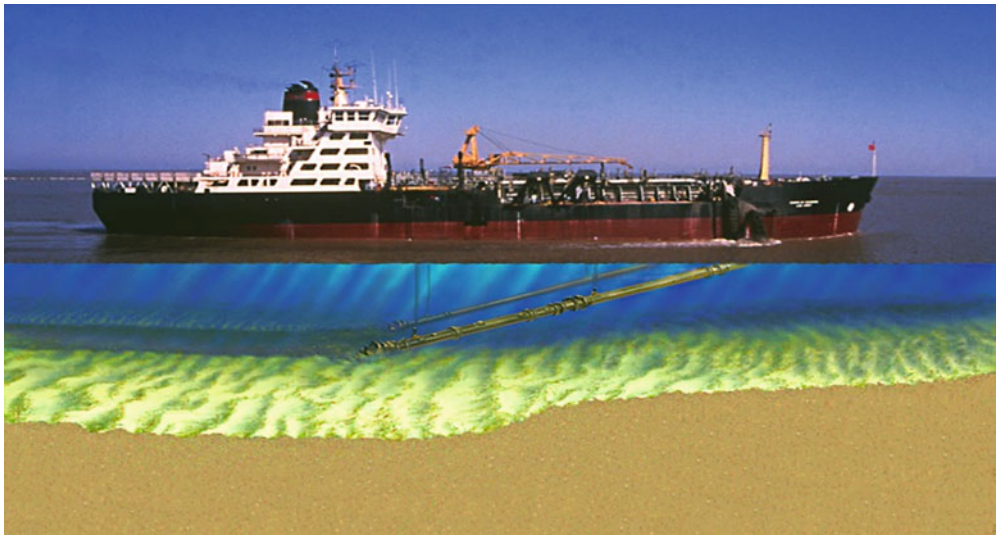
The centrifugal pump on the hydraulic dredge operates at or near the water surface elevation to pull water into the suction mouth at the cutterhead. The combination of the vacuum and atmospheric weight acts to move the bed material up through the suction pipe to the pump, and then the pump discharges the slurry into the discharge pipeline and directly to the disposal site. Because pipeline dredges pump directly to the disposal site, they operate continuously and can be very cost efficient. A booster pump is used for long distances to the disposal site (Fig. 6).

Cutterhead pipeline dredges work efficiently in large areas and in water depths up to 60–70 ft. The dredge production rate is the most efficient and creates minimum excess turbidity at the cutterhead when the loose sediment cut depth is equal or near the diameter of the dredge cutterhead. The dredged material slurry typically contains 90% water and 10% solids volume, and the water must be contained in the disposal site until the solids settle out. The dewatering is accomplished typically by discharging surface water in the disposal site back into the waterway. If it is contaminated sediment, discharge controls are required to minimize impacts to the receiving waters.

The discharge pipe can be a floating pipe on the surface or a submerged discharge pipe. Typically, the discharge line is a floating line, and they are not well suited for work in rough seas, where lines can be broken apart, or in high traffic areas, where the discharge pipeline can be an obstruction to navigation. The pipeline dredge is not an efficient dredge plant to work in an open estuary, entrance channel, or other open-water areas where significant wave conditions may occur. Rough waters with wave heights of significance will cause damage to the cutterhead, spuds, and the dredge hull, as well as the floating pipeline.

Hopper dredges are ships designed for dredging (Figs. 7 and 8). The trailing suction hopper dredge is

Dredging Practices and Environmental Considerations. Figure 6
Pipeline booster pump (Courtesy: Great Lakes Dredge & Dock)



Dredging Practices and Environmental Considerations. Figure 7
Typical hopper dredge (Photo courtesy Corps of Engineers)

a self-propelled seagoing ship equipped with a suction pipe, which trails over the side of the vessel or through a well in the hull. The sediment and water slurry is transported through the pumps just as the pipeline dredge, but when the sediment and water slurry passes through the pump to the discharge pipeline, it is discharged immediately into the hoppers of the dredge. When the hoppers are full, the sediment

Dredging Practices and Environmental Considerations. Figure 8
Hopper dredge, Liberty Island (Photo courtesy of Great Lakes Dredge & Dock Company)

and water slurry is transported by the ship to the disposal site.

A hopper dredge can dispose of the dredged sediment in two manners. The split-hull hopper dredge design and the standard hopper dredge with bottom dump capability will discharge the dredged sediment to a submerged and acceptable disposal site. The alternative to bottom dumping is to pump out of the hoppers to an upland disposal site or to a shallow water area that the vessel draft prevents access.

The hydraulic hopper dredge is generally comprised of the following equipment:

- Drag arm. The drag arm on the hopper dredge serves the same purpose as the ladder on the pipeline dredge.
- Drag head. The drag head on the hopper dredge provides a similar purpose as the cutterhead on the pipeline dredge. Several designs of drag heads are used for different sediment types. The drag head design includes:
  - Erosional drag heads (unconsolidated fine sediment, sand, and gravel).
  - Mechanical drag head (silts, mud, light clay).
  - Combination drag head with waterjets (hard-packed sand to gravel).
- Gimbal joints. The gimbal joints allow the drag arm to articulate while the vessel moves, thereby keeping the drag head on the bed during mild to medium sea conditions.
- Swell compensator. The swell compensator acts much like a shock absorber and allows the hopper dredge to keep the drag heads on the bed and still work in relatively high-wave conditions (8–10 ft).

The slurry discharge is into the hoppers of the ship. Hopper dredge size, in terms of dredging, is defined by the capacity of the hoppers. The slurry of water and sediment is discharged into the top of the hopper. The sediments settle out in the hopper, while excess water is separated and discharged over the weir structure. It can be advantageous to overflow hopper dredges to increase the amount of sediment in the hopper; however, this is not always acceptable due to water quality concerns near the dredging site.

Hopper dredges are well suited to dredging heavy sands. They can maintain operations in relatively rough seas, and because they are mobile, they can be used in high-traffic areas. They are often used at ocean entrances where wave conditions prevent use of pipeline dredges and limit mechanical dredging capability. They cannot be used efficiently in confined or shallow

areas. Hopper dredges can move quickly to disposal sites under their own power. However, because the actual removal of sediment from the bed stops during the transit to and from the disposal area, the operation loses efficiency that can become critical when the haul distance to disposal is far.

There is a special hopper dredge class called side-casters and pipeline dredge called dustpan dredges. Both of these dredges are unique within their family of dredges. They are specialty dredges used for unusual waterway conditions. They work best when dredging loosely compacted, coarse-grained, clean sediment with disposal in areas close to the dredging activity. They are not widely used.

Environmental Cleanup Dredges

Dredging of contaminated sediments is potentially very harmful to the local environment during dredging and disposal. Contaminants can be remobilized and/or released into the water column where they can detrimentally affect aquatic life and pose a risk to human health. Technological advances have fostered modification of existing dredge equipment, and creation of new dredging equipment to address the environmental issues. Contaminated sediment dredging focuses on minimizing suspension and release of problem sediments in the water column while increasing the precision of dredging to reduce overdredging. Examples of contaminated sediment dredges include the following:

- Encapsulated bucket lines for bucket chain dredges.
- Closed buckets for backhoes.
- Closed clamshells for grab dredges.
- Auger dredges, disk cutter, scoop dredges, and sweep dredges (all modified cutter dredges) [4].

Transportation of Dredged Material

Transportation methods generally used to move clean and contaminated dredged materials are included in the three basic dredge types: pipelines, barges or scows, and hopper dredges.

- Pipeline transport is the method most commonly associated with cutterhead, dustpan, auger head, and other hydraulic dredges. Dredged material may be directly transported by hydraulic dredges through pipelines for distances of up to several miles, depending on a number of conditions. Longer pipeline pumping distances are feasible with the addition of booster pumps, but the cost of transport greatly increases proportionally with each booster pump added to the discharge line.
- Barges and scows, used in conjunction with mechanical dredges, have been one of the most widely applied methods of transporting large quantities of dredged material over long distances.
- Hopper dredges are capable of transporting the material for long distances in a self-contained hopper. Hopper dredges normally discharge the material from the bottom of the vessel hull by opening the hopper doors; however, most hopper dredges are equipped to pump out the material from the hopper and deliver the sediment much like a hydraulic pipeline dredge.

Dredged Material Disposal and Placement Alternatives

Evaluation and design of a proposed dredging project involves comprehensive assessment of alternatives for disposal or placement of the dredged material. Identification of the specific disposal site or beneficial use site involves a number of different considerations, including environmental, technical, and economic factors.

Three major disposal/placement alternatives are available:

- Open-water disposal in deep waters or along banks of a river outside the navigation channel
- Confined disposal in open water (confined aquatic disposal (CAD)) and on land (confined disposal facility (CDF))
- Placement for environmental and beneficial use

In the case of very contaminated sediments and cleanup/remedial dredging, treatment of the dredged material after temporary storage and before final disposal may be a necessary alternative.

Open-Water Disposal

Open-water disposal means that dredged material is placed at designated sites in oceans, estuaries, rivers, and lakes such that it is not isolated from the adjacent water. Clean dredged materials are the only acceptable dredged materials for disposal at open-water disposal

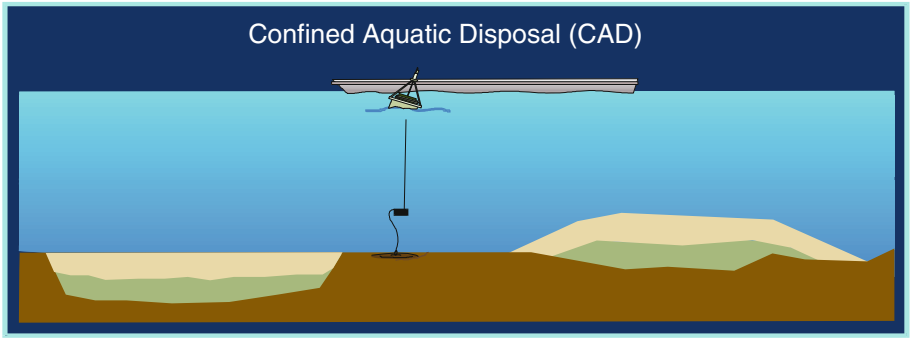
sites. The determination that dredged material is "clean" is based upon a series of chemical and biological tests, the results of which must meet national environmental regulations. The disposal of contaminated material can be considered for open-water disposal but only with appropriate control measures, such as capping the contaminated sediments by the use of clean capping materials.

The objective of capped in-water disposal is to isolate contaminated materials from the environment by covering the contaminated materials with clean materials, such as fine to coarse sand. The contaminated material is placed on a level bottom, in engineered deep constructed pits, or in bottom depressions. The cap of clean sediment that is placed on top must be designed to withstand erosion over time from bottom currents, waves, vessel movement, and prop wash, and burrowing bottom creatures (Fig. 9). Caps should be monitored over time to ensure their integrity [5].

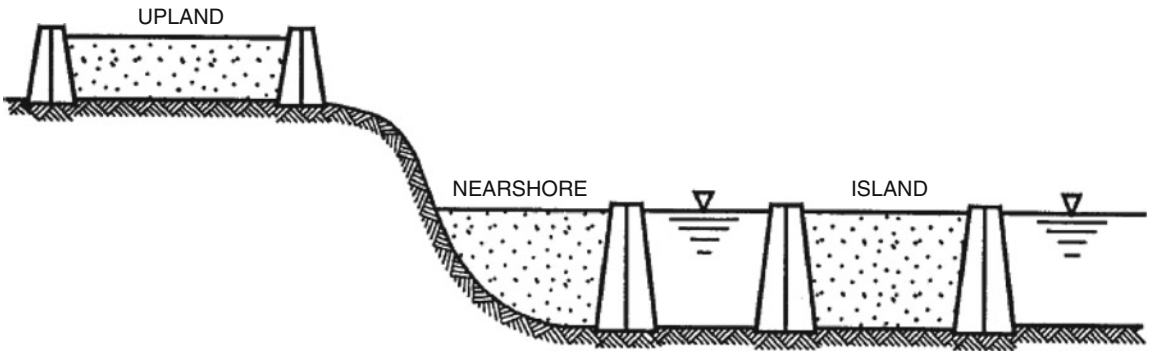
Confined Disposal Facilities

Confined disposal is placement of dredged material within engineered diked nearshore or upland confined disposal facilities (CDFs) via pipeline or barge delivery of sediments. CDFs may be constructed as upland sites, nearshore sites with one or more sides in water (sometimes called intertidal sites), as island containment areas, or as subaqueous contained capped cells (Figs. 10–12).

The two objectives inherent in design and operation of CDFs are to provide for adequate storage capacity to meet dredged volume requirements and to maximize efficiency in retaining the solids. For facilities

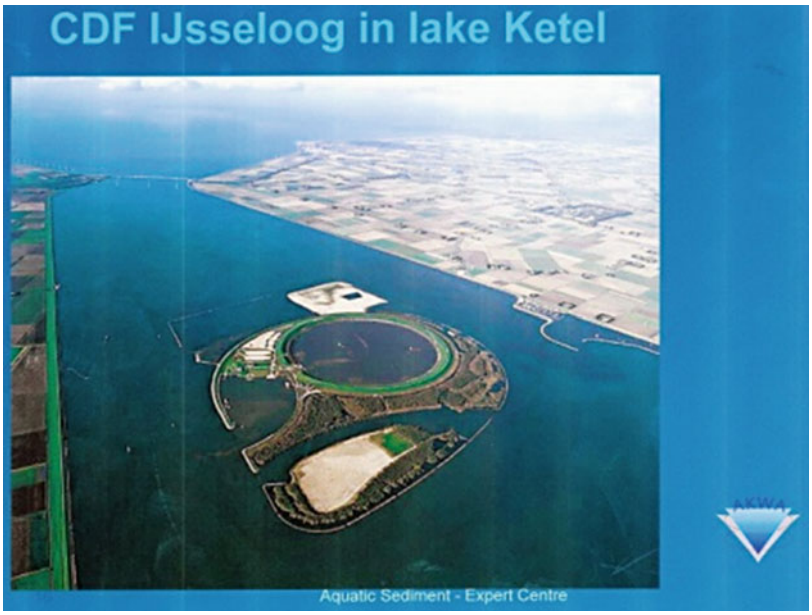


Dredging Practices and Environmental Considerations. Figure 9
Confined aquatic disposal (CAD) (Courtesy of Corps of Engineers)



Dredging Practices and Environmental Considerations. Figure 10
Types of confined disposal facilities (Courtesy of Corps of Engineers)

Dredging Practices and Environmental Considerations. Figure 11
Nearshore CDF Huelva Estuary, Spain (Courtesy Spain government)



Dredging Practices and Environmental Considerations. Figure 12
Island CDF at IJsseloog, the Netherlands (Courtesy Dutch government)

receiving contaminated material, an additional objective is to provide the efficient isolation of contaminants from the surrounding area. To achieve these objectives, depending on the degree of intended isolation, CDFs may be equipped with a complex system of control measures, such as surface covers and bottom liners, treatment of effluent, surface runoff and leachate monitoring, and management controls.

Hydraulic dredging adds several volumes of water for each volume of sediment removed. This excess water is normally discharged as effluent from the CDF during the filling operation. The amount of water added depends on the design of the dredge, physical characteristics of the sediment, and operational factors, such as the pumping distance. When the dredged material is initially removed from the bed and deposited in the CDF, it can and will occupy several times its original volume. The settling process is a function of time, with sandy and gravelly sediment dewatering quickly and silts and clays dewatering very slowly. Silts and clays will eventually consolidate to the loose in situ volume or less if desiccation occurs. Adequate volume must be provided during the dredging operation to contain the total volume of sediment to be dredged, accounting for any volume changes during placement.

Beneficial Use of Dredged Material

Dredged material is increasingly regarded as a resource rather than as a waste. More than 90–95% of sediments from navigation dredging comprise sediments acceptable for open-water disposal; these are also considered acceptable for a wide range of environmentally and economically beneficial uses. The first step in examining dredged material management options is to consider possible beneficial uses of dredged material. Recent decades have seen the increasing use of dredged materials for habitat creation, habitat restoration, beach nourishment, and coastal protection (Fig. 13).

Beneficial use is defined as "*Utilizing dredged sediments as resource materials in productive ways, which provide environmental, economic, or social benefits*" [6]. Broad categories of beneficial uses of dredged material, based on the functional use of the dredged material or site, include:

- Habitat development and restoration
- Parks and recreation
- Coastal protection
  - Beach nourishment
  - Riverbank and lakeshore protection
- Nearshore placement/littoral zone sediment management
- Construction and agricultural
  - Construction and industrial/commercial development (roads, dikes, levees, parking lots)
  - Land reclamation/remediation (brownfield restoration, strip mine reclamation)
  - Agriculture, forestry, horticulture, and aquaculture

Dredging Practices and Environmental Considerations. Figure 13
Deer Island Marsh Creation: Mississippi, USA (Courtesy of Corps of Engineers)

Dredging Practices and Environmental Considerations. Figure 14
Beneficial use site: Evia Island Bird Nesting, Galveston Bay, Texas (Courtesy Port of Houston Authority)

Operational feasibility for these projects includes the availability of suitable material in the required amount at a particular time and within the project costs. These are crucial aspects of the most beneficial uses projects (Fig. 14).

Treatment of Dredged Material

In certain cases of environmental cleanup by dredging, treatment of the dredged material may be necessary prior to confined disposal or reuse. A variety of treatment technologies are available to reduce the quantity or to reduce the contamination of the dredged material.

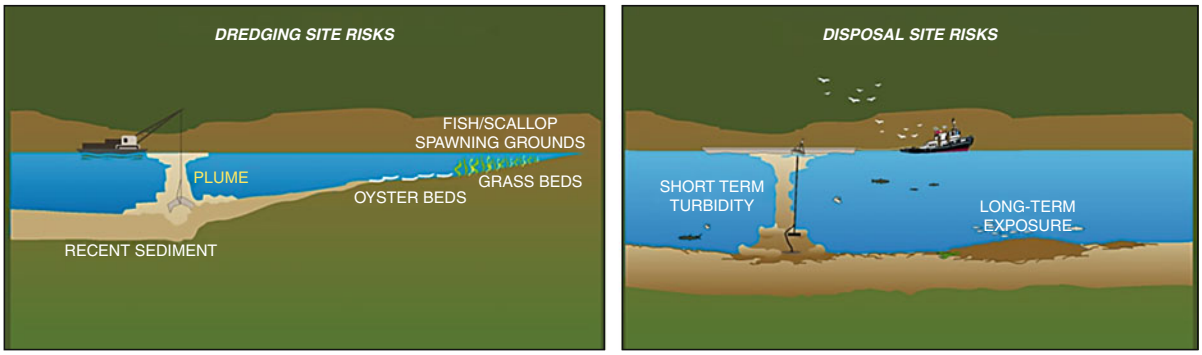
Treatment methods range from separation techniques, in which contaminated sediments are separated from relatively clean sand, to incineration. Some techniques are well developed to date, but others are still in the early stages of development. The products of the treatment methods have a wide array of potential beneficial uses (e.g., construction grade cement, lightweight aggregate, bricks, and manufactured soil). The problem is that treatment is usually very expensive, thereby limiting project scale. As a result, the treatment of small volumes of contaminated material is more likely than that of large volumes.

Environmental Considerations

The potential environmental effects of navigation dredging and environmental cleanup dredging are the result of the actual dredging activity in the water and a result of the disposal of the dredged material (Fig. 15).

During dredging, effects may arise due to the excavation of sediments causing resuspension in the water column, loss of material during transfer to the barge, overflow from the dredge while loading and loss of material from the hopper dredge and/or pipelines during transport to disposal. Potential effects during disposal arise depend upon the physical and chemical characteristics of the dredged material and the selected disposal site (i.e., open-water, nearshore, or upland).

During all dredging operations, the removal of material from the seabed also removes the surface-based (benthic) animals living on and in the sediments (benthic animals). With the exception of deep-burrowing animals or mobile surface animals that may survive a dredging event through avoidance, dredging can initially result in the complete removal of surface-dwelling biota from the dredging site. Where the channel or berth has been subjected to regular maintenance dredging over many years, it is very unlikely that well-developed benthic communities will occur in or around



Dredging Practices and Environmental Considerations. Figure 15
Environmental risks: At the dredging and disposal sites (Courtesy Corps of Engineers)

the dredged area. The recovery of disturbed habitats following dredging ultimately depends upon the nature of the new sediment at the dredge site, sources and types of recolonizing benthos, river width and bank line, and the extent of the disturbance [7].

The environmental issues associated with dredging and dredged material management include:

- Physical and ecological impacts due to turbidity and sedimentation.
- Ecological and human health impacts: acute and chronic toxicity due to chemical contamination, e.g., PCBs, PAHs, dioxin, metals, such as lead, cadmium, and mercury.
- Loss of habitat – due to dredging or placement of dredged material.
- Impacts to endangered species (e.g., turtles) due to dredging.
- Impacts to fish migration and spawning due to turbidity and exposure to toxic chemicals from dredging and disposal.
- Impacts of noise from dredging operations upon aquatic living resources.
- Others: emissions of air pollutants, quality-of-life issues (e.g., noise, night lights).

When dredging and disposing of noncontaminated fine materials (e.g., silts, clays) in estuaries and coastal waters, the main environmental effects are associated with suspended sediments and increases in turbidity. All methods of dredging release suspended sediments into the water column during the excavation itself and during the overflow of dredging water from hoppers and barges. In many cases, the locally increased suspended sediments and turbidity associated with dredging and disposal are obvious from the turbidity "plumes," which may be seen trailing behind dredges and disposal sites [8].

- Increases in suspended sediments and turbidity levels from dredging and disposal operations may under certain conditions have adverse effects on marine animals and plants by reducing light penetration into the water column and by physical disturbance.
- Increases in suspended sediments can impact filter-feeding organisms, such as shellfish, through clogging and damaging feeding and breathing equipment. Similarly, young fish can be damaged if suspended sediments become trapped in their gills, and increased fatalities of young fish have been observed in heavily turbid waters. Adult fish are likely to move away from or avoid areas of short-term high suspended solids, such as dredging sites, unless food supplies are increased as a result of increases in organic material.
- In important spawning or nursery areas for fish and other marine animals, dredging can result in smothering eggs and larvae. Shellfish are particularly susceptible during the spring when spatfall occurs.
- Increases in turbidity result in a decrease in the depth that light is able to penetrate the water

column which may affect submerged seaweeds and plants, such as eelgrass *Zostera* species, by temporarily reducing productivity and growth rates.

The degree of resuspension of sediments and turbidity from dredging and disposal depends on four main variables:

- The sediments being dredged (size, density, and quality of the material).
- Method of dredging (and disposal).
- Hydrodynamic regime in the dredging and disposal area (current direction and speed, mixing rate, tidal state).
- The existing water quality and characteristics (background suspended sediment and turbidity levels).

The resuspension of sediments during dredging and disposal may also result in an increase in the levels of organic matter and nutrients available to marine organisms. In certain cases, nutrient enrichment can lead to the formation of algal blooms (eutrophication). These blooms can reduce the surrounding water quality by causing the removal of oxygen as the blooms break down, or occasionally, by the release of toxins, which may disturb marine wildlife.

Sediments dispersed during dredging and disposal may resettle over the seabed and the animals and plants that live on and within it. This blanketing can cause smothering of benthic animals and plants, may cause mortality, stress, and reduced rates of growth or reproduction, and, in the worst cases, the effects may be fatal. Generally, sediments settle within the vicinity of the dredged area, where they are likely to have little effect on the recently disturbed communities, particularly in areas where dredging is a well-established activity. However, in some cases, sediments are distributed more widely within the estuary or coastal area and may settle over adjacent subtidal or intertidal habitats possibly some distance from the dredged area.

When dredged materials are placed in open-water disposal sites, they will have a blanketing and smothering effect on benthic organisms in the immediate disposal site. The continual disposal of maintenance dredging at disposal sites may prevent the development of stable benthic communities and the partial or complete loss of benthic production and habitat. Recolonization is expected when disposal operations have been completed, depending on the characteristics of the dredged material and the changes to the hydrodynamic conditions at the disposal site.

A variety of harmful substances, including heavy metals, tributyl tin (TBT), polychlorinated biphenyls (PCBs), and pesticides, are in the sediments in certain ports, harbors, and waterways. These contaminants are often of historic origin and from local or upstream sources. The highest levels of contaminants generally occur in industrialized estuaries. Dredging and disposal can release these contaminants into the water column, making them available to be taken up by animals and plants, with the potential to cause adverse acute and chronic toxicity. The risk of this occurring depends upon the type and degree of sediment contamination. If contaminants are released into the water column or are in the sediments at the open-water disposal site, they may bioaccumulate in marine animals and plants and transfer up the food chain to fish and sea mammals, with associated risks to human health.

Dredging can cause direct threats to endangered species, such as sea turtles and their nearshore marine habitats. Hopper dredges have been directly responsible for the incidental capture and the death of hundreds, if not thousands, of sea turtles in the United States. Development of specially designed hopper dredge drag heads and institution of best management practices in areas of turtle populations has helped alleviate the majority of the takings of turtles during dredging operations [9].

Nearshore or upland CDFs are the most commonly used disposal technique for contaminated dredged material. Pathways for potential exposure to animals/plants and humans are similar for these two types of disposal sites. Comparison to open-water sites is not necessarily appropriate since open-water sites should be receiving clean dredged material, while nearshore and upland sites isolate contaminated dredged material from the surrounding environment. Potential pathways include the discharge into receiving waters (e.g., estuary or river) of the excess water from the dredged material, contamination of ground water, and exposure of birds and animals to the dredged material in the CDF. Depending upon the level of contamination, controls can be used to minimize

negative environmental impacts, such as using of impervious liners for disposal sites receiving dredged material from cleanup dredging.

Environmental Regulation of Dredging and Dredged Material Disposal/Placement

In addition to national and regional legislation and policies, the most widely applicable international regulatory instrument is the London Convention 1972 and London Protocol 1996 (LC/LP), which is part of the International Maritime Organization, an organization of the United Nations. The LC/LP regulates disposal of wastes into ocean waters, worldwide [10]. The LC/LP is an international treaty, which includes 90 country signatories. Member countries are required to implement the conditions of the treaty including the waste assessment procedures noted below.

The LC/LP Waste Assessment Guidelines for Dredged Material allow disposal of dredged material into ocean waters, provided that strict environmentally protective criteria are met. A step by step process to evaluate a dredging project, the alternatives for disposal or placement for beneficial use, and an action list for judging environmental acceptability of open-water disposal are specified in the Guidelines [11]. Components of the guidelines are shown in Fig. 16.

After an assessment of the need for dredging, major dredging or disposal projects should have studies carried out in order to ensure that any potential adverse effects are identified in advance and dealt with in an appropriate manner. Such investigations include characterization of the dredged material (physical, chemical, and toxicity), an examination of any sources of contamination and the potential to control those sources, an assessment of disposal or beneficial use placement alternatives, including identification and characteristics of the disposal site, and design of monitoring studies to determine whether any potential impacts are correctly predicted.

The environmental impact assessment should highlight both positive and negative, short- and long-term impacts. Appropriate testing may be required to determine the physical behavior of the material at the disposal site. Also, testing and assessments of potential contaminants of concern may be required, depending upon existing knowledge of the dredging site and any potential contaminant pathways. Where potentially adverse effects are anticipated, management techniques should be implemented to reduce risks to acceptable levels. Possible controls for open-water alternatives, include operational modifications, use of submerged discharges of dredged material, treatment, lateral containment, and capping or contained aquatic disposal. Possible controls for confined disposal facilities, include operational modifications, treatment, and various site controls (e.g., covers and liners).

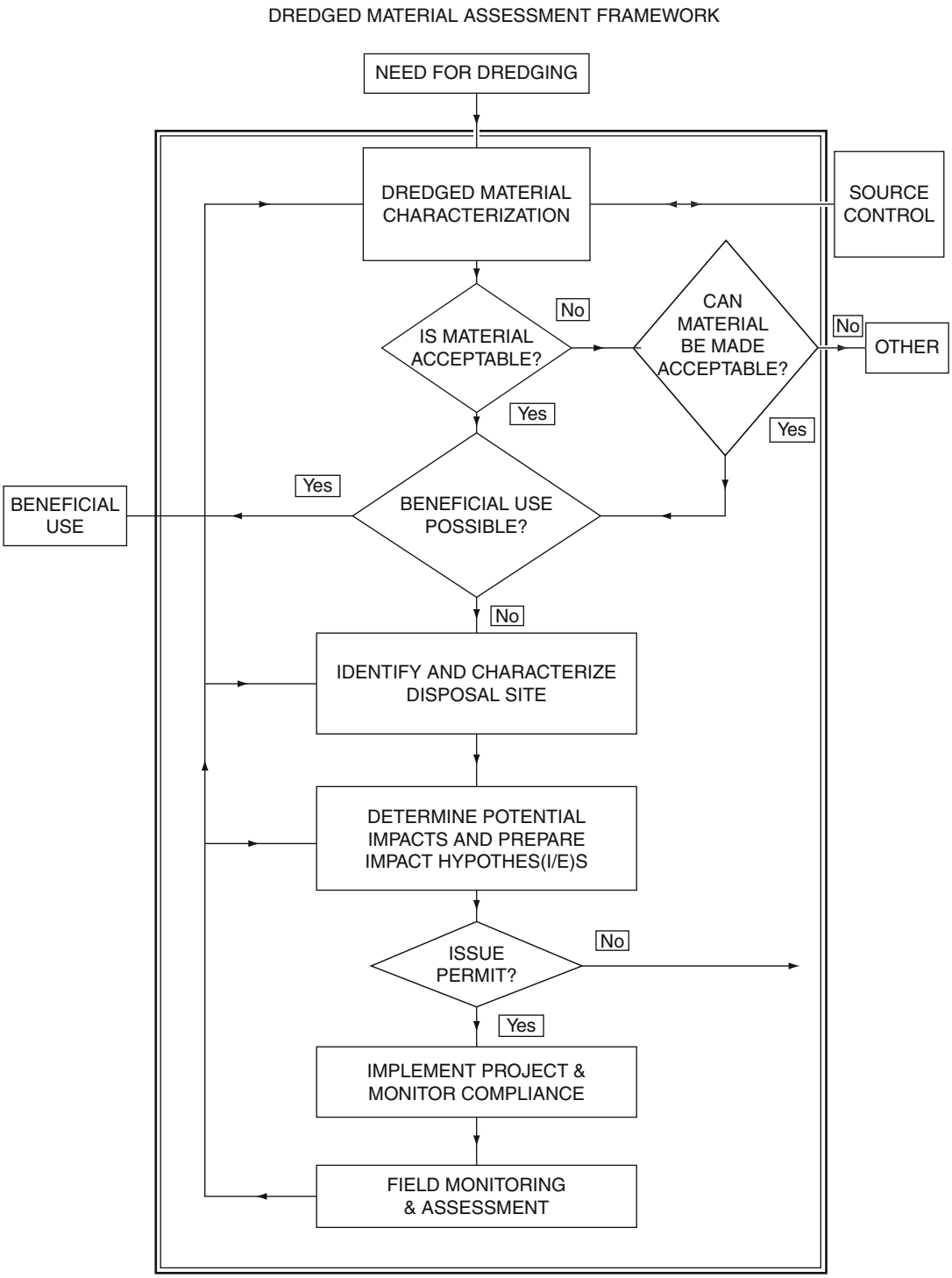
An important component in development of the environmental impact assessment and in identifying potential impacts and implementing acceptable measures is the involvement of interested groups and organizations, consulting with them, and reaching a consensus in the early in the process of determining the alternatives. It is in the best interests of the project sponsors and stakeholders that the decision-making process is transparent, stakeholders are involved, and that the reasons for the selection of the preferred dredging and disposal or placement options are clearly understood.

Control of Upstream Sources of Sediments and Contaminants

Control of upstream sources of sediments and contaminants represents a significant part of the long-term solution to the continuing need for maintenance dredging of navigation channels and the continuing issues of contaminated dredged material. While control of upstream sources of sediments can contribute to lessening the need for dredging, sediments play a duel role in the watershed:

1. Sediments cause water quality problems and the need for more frequent dredging.
2. Sediments provide shoreline protection and maintaining habitats, such as marshes.

Sediments are increasingly recognized as scarce resources that can have costly impacts if not carefully managed, recycled, and conserved. Sediment overloading from land and stream erosion causes significant environmental and economic challenges, i.e., excessive needs for navigation dredging, and sediment in rivers, reservoirs, and estuaries may contribute to high turbidity, which can affect both habitat for aquatic life and human water uses, loss of flood storage



Dredging Practices and Environmental Considerations. Figure 16
International guidelines for assessment of dredged material (Courtesy London Convention)

capacity, and obstructions to waterway navigation and conveyance of floodwaters. Yet in other locations, a shortage of sediment can result in coastal and streambank erosion, and wetland loss. Many water resource projects are designed to remedy local sediment problems, but they must be well planned and thought out as they can create even larger problems upstream or downstream of the local problem.

Dredging and other projects affecting or involving sediment should be undertaken with awareness of the littoral or fluvial sediment system and the effects that proposed projects and actions may have on other stakeholders. Regulatory authorities and stakeholders are becoming increasingly aware of the necessity to manage sediments on a system-wide basis, dredging being one part of the overall river, lake, or coastal system [12].

Principles of Sustainable Sediment Management

- Recognize sediment as part of a system and as a valuable resource that is integral to economic and environmental vitality.
- Strive for balanced, economic, and environmentally sustainable solutions to sediment-related issues through integrated management of sediment from upland sources to estuaries and within coastal zones.
- Involve external and internal partners and stakeholders to integrate and balance objectives and to leverage resources in implementation.
- In project decision-making, consider the sediment implications beyond the local site, including intended and potential effects, and over long time scales (decades or more), and be effective stewards of sediment and related resources.

Dredging regulatory authorities and dredging sponsors need to embrace the concept of sediment management within the watershed. They should work with upstream pollution control authorities to address sediment management decisions such that the overall sediment system is evaluated in dredging projects. Similarly, upstream sources of contaminants from industrial and municipal sources as well as diffuse sources, such as storm runoff from farms or urban areas, need to be understood and comprehensive controls developed.

Achieving sediment management for source control usually involves substantial collaboration and cooperation between a number of organizations, agencies, and stakeholders. These groups must be willing to identify the source and take measures to reduce or prevent further contamination of waterways with sediments and chemicals. This is no easy task but it is the only viable way to guarantee a long-term successful outcome of our waterways development and environmental protection [13].

Future Directions: Sustainable Dredging and Dredged Material Management

The global economy and the dependence upon food and commodities via international trade require that vessels have sufficiently deep channels in ports, harbors, and waterways for safe passage. Other interests include national security as well as recreational opportunities.

While upstream sediment management controls will help, the natural erosion process in rivers and estuaries will continue. Thus, navigation dredging will continue to be needed over the very long term. Environmental cleanup dredging will be needed for decades to come, even as improved controls are placed upon waste and wastewater sources. Legacy contaminants already in the sediments will continue to pose aquatic and human health risks until they are removed or isolated from the surrounding aquatic environment. The environmental considerations relate to the quality and quantity of the sediment to be dredged, the potential environmental risks from the dredging itself, and what to do with the dredged material.

Over time, dredging and dredged material management practices will move towards sustainability concepts. The likely trends include:

Sediment Management

Dredging projects will be managed as part of the overall sediment system in the watershed, with dredged material being considered a resource. Opportunities for beneficial use of dredged material will increase as potential beneficial use projects (e.g., habitat restoration or creation, beach nourishment and coastal protection, restoration of brownfields, and construction

purposes) and their sponsors are identified early in the dredging project planning process.

Climate change can cause more erosion in some places and less in others, with associated changes in the quantities of sediment needed to be removed by dredging for navigation purposes. Continued focus upon control of local and upstream sources of sediment and contaminants will begin to pay dividends by reducing the frequency of navigation dredging. Improved chemical and toxicological quality of those sediments will occur as additional environmental controls are put in place to control municipal and industrial discharges and storm water runoff from urban and rural areas, including farmlands. Instituting environmental controls on the disposal of hazardous waste has helped control runoff from hazardous waste disposal sites and over decades will reduce the need for environmental cleanup dredging. Cleanup dredging, such as removal of PCB-contaminated sediments in the Hudson River, New York; Fox River, Wisconsin; and Passaic River, New Jersey, will contribute to improved sediment quality downstream.

Technological Innovations

Driven by concerns about the potential impacts to aquatic life and human health, technology will continue to evolve in dredging hardware, treatment of contaminated dredged material, and use of dredged material in beneficial use applications. Innovations in dredging technology are focused upon reduction in the disbursement of suspended solids and associated contaminants into the water column during dredging and disposal operations.

Further technological innovations in the types and efficiencies of treatment technologies will likely identify potential reuse opportunities for certain dredged materials, such as use as soil, fill, or aggregates. While the science and engineering of confined disposal facility design is well established, the likely trend is to increase use of confined aquatic disposal cells, ensuring the isolation of contaminated sediment from the aquatic environment.

Two other areas of dredging and disposal are likely to see significant changes:

1. Electric powered dredges will contribute fewer diesel emissions and NO_x in locations where compliance with air pollution standards is an issue or regulation.
2. Changes in navigation channel design to accommodate larger ships (deeper channels) will impact dredging projects, and improved channel design will be necessary due to limited project funding (e.g., narrower channels and institution of vessel operational controls, and fewer deep water ports with attendant increases in the use shallower draft vessels to move cargo between coastal ports).

Implementation of Regulations

International guidelines (i.e., London Convention/London Protocol) are in place for protection of the environment from dredged material disposal in ocean waters. National and local regulations are in place in many countries that implement the London Convention/London Protocol guidelines as well as for protection of internal country waters. These regulations are in various stages of implementation worldwide. Technical cooperation and assistance programs are ongoing to assist developing countries in their application.

One key aspect of the national and local regulations is the characterization of the dredged material prior to disposal. Updated procedures for testing dredged material will provide better techniques to assess its acceptability for open-water disposal or for specific beneficial uses; these include improved bioassays, interpretive guidance, and the application of risk assessment in cases where high uncertainties exist [14].

The final trend in the regulatory arena is the use of mitigation for unavoidable adverse impacts due to dredging and disposal. This will become more widespread as one of the tools for regulatory authorities.

Bibliography

Primary Literature

1. Willard (2009) *Willardsays. . ..*Dredge history done lite. http://www.willardsays.com/dredge_history.pdf. Accessed 25 May 2011
2. Cohen M (2005) Dredging: the environmental facts. Published by International Association of Dredging Companies; Typeset and printed by Opmeer Drukkerij bv, The Netherlands ISBN 90-75254-11-3. http://www.dredging.org/documents/ceda/downloads/publications-dredging_the_facts. Accessed 1 April 2011

3. U.S. Army Corps of Engineers, U.S. Environmental Protection Agency (2004) Evaluating environmental effects of dredged material management alternatives – A technical framework. EPA842-B-92-008
4. Bray RN, International Association of Dredging Contractors, Central Dredging Association (2008) Environmental aspects of dredging. Taylor and Francis/Balkema, Amsterdam
5. Otten MT, Hartman G (2002) Placing and capping fine-grained dredged material. In: ASCE conference proceedings of the third specialty conference on dredging and dredged material disposal, May 5–8, Orlando, Florida, 119
6. USEPA, Corps of Engineers (2007) Identifying, planning, and financing beneficial use projects using dredged material beneficial use planning manual. U.S. EPA, EPA842-B-07-001, Corps of Engineers, Washington, DC, USA
7. International Council for the Exploration of the Sea (ICES) (1992) Report of the ICES working group on the effects of extraction of marine sediments on fisheries. ICES cooperative research report # 182, Copenhagen (Denmark)
8. UK Marine Protected Areas Centre (2001) UK marine SACs project. Dredging and disposal. http://www.ukmarinesac.org.uk/activities/ports/ph5.htm. Accessed 2 April 2011
9. U.S. Corps of Engineers, Sea Turtle Data Warehouse. http://el.erdc.usace.army.mil/seaturtles/intro.cfm. Accessed 27 September 2011
10. IMO (2003) London convention 1972 and the 1996 protocol, 2nd edn. International Maritime Organization, London
11. IMO (2007) London convention 1972. Specific guidelines for assessment of dredged material
12. U.S. Army Corps of Engineers (2004) Regional sediment management – primer. U.S. Army Corps of Engineers, Washington, DC
13. Vogt C (2007) Dredged material management in a watershed context – seeking integrated solutions. In: Proceedings of 26th WEDA conference, San Diego
14. Moore DW, Bridges TS, Ruiz C, Cura J, Driscoll SK, Vorhees D, Peddicord D (1998) Environmental risk assessment and dredged material management: issues and application. In: Proceedings of workshop, San Diego

Books and Reviews

Adams JR, Herbich JB (1970) Gas removal systems. In: Proceeding world dredging conference, WODCON'70, Singapore

Addie GR, Whitlock L, Dresser J (2001) Dredge pump testing considerations. In: Proceedings of the Western Dredging Association 21st technical conference and 33rd Texas A & M Dredging seminar, Houston, TX, 24–27 Jun 2001

Albar A (2001) Modeling of a bucket wheel dredge system for offshore sand and tin mining. Ph.D. Dissertation, Ocean engineering program, Texas A & M University, Ocean engineering program, Civil Engineering Department, College Station, TX.

Ancheta C, Santiago R, Sherbin G (1998) An innovative approach to harbour remediation, Thunder Bay, Ontario, Canada. In: Proceedings of 15th world dredging congress, Las Vegas, NV

Board M (1997) Contaminated sediments in ports and waterways. National Research Council, National Academy Press, Washington, DC

Bray RN, Bates AD, Land JM (1997) Dredging: a handbook for engineers, 2nd edn. Arnold, London

Brandon DL, Price RA (2007). Summary of available guidance and best practices for determining suitability of dredged material for beneficial uses. ERDC/EL TR-07-27. U.S. Army Engineer Research and Development Center, Vicksburg

Bridges TS, Ells S, Hayes D, Mount D, Nadeau SC, Palermo MR, Patmont C, Schroeder P (2008) The four Rs of environmental dredging: resuspension, release, residual, and risk, ERDC/EL TR-08-4. U.S. Army Engineer Research and Development Center, Vicksburg

Briot JE, Doody JP, Romagnoli R, Miller MM (1999) Environmental dredging case studies: a look behind the numbers. In: Proceedings of the Western Dredging Association 19th technical conference and 31st Texas A & M Dredging seminar, Louisville, KY, 15–18 May 1999

Cura JJ, Heiger-Bernays W, Bridges TS, Moore DW (1999) Ecological and human health risk assessment guidance for aquatic environments. U.S. Army Corps of Engineers, Engineer Research and Development Center, Vicksburg

Demars KR (1995) Dredging, remediation, and containment of contaminated sediments. ASTM committee D-18 on soil and rock, environment, Canada

Dickerson DD, Nelson DA, Dickerson CE Jr, Reine KJ (1993) Dredging related sea turtle studies along the Southeastern U.S., Coastal Zone 93, New Orleans, LA

Eisma D (2005) Dredging in coastal waters. Taylor & Francis, London

Fredette TJ, Anderson G, Payne BS, Lunz JD (1986) Biological monitoring of open-water dredged material disposal sites. In: IEEE Oceans'86 conference proceedings, Washington, DC

Fredette TJ, Nelson DA, Clausner JE, Anders FJ (1990a) Guidelines for physical and biological monitoring of aquatic dredged material disposal sites, Technical report D-90-12. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Fredette TJ, Nelson DA, Miller-Way T, Adair JA, Sotler VA, Clausner JE, Hands EB, Anders FJ (1990b) Selected tools and techniques for physical and biological monitoring of aquatic dredged material disposal sites, Technical report D-90-11. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Graalum SJ, Randall RE, Edge BL (1999) Methodology for manufacturing topsoil using sediment dredged from the Texas Gulf intracoastal waterway. J Marine Environ Eng 1:1–38

Hayes DF (1986) Guide to selecting a dredge for minimizing resuspension of sediment, Environmental effects of dredging, Technical notes, EEDP-09-1. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Hayes DF (1999) Turbidity Monitoring during boston harbor bucket dredge comparison study. Project report, Submitted to USAE Waterways Experiment Station, Vicksburg

Hayes DF, McLellan TN, Truitt CL (1988) Demonstration of innovative and conventional dredging equipment at Calumet

Harbor, Illinois, MP EL-88-01. U. S. Army Engineer Waterways Experiment Station, Vicksburg

Hayes DF, Raymond GL, McLellan TN (1984) Sediment resuspension from dredging activities. In: Dredging 1984. American Society of Civil Engineers, Clearwater

Herbich JB, Brahme SB (1991) A literature review and technical evaluation of sediment resuspension during dredging, Contract report HL-91-1. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Herbich JB (2000) Dredge pumps. In: Herbich JB (ed) Handbook of dredging engineering. McGraw-Hill, New York, pp 3.36–3.58, Chapter 3

Herbich JB (2000) Removal of contaminated sediment by dredging. In: Dredging engineering short course notes, Center for Dredging Studies, Ocean Engineering Program, Civil Engineering Department, Texas A & M University, College Station

Herbich JB (2000) Handbook of dredging engineering, 2nd edn. McGraw-Hill, New York

Herbich JB (1995) Removal of contaminated sediments: equipment and recent field studies. In: Demars KR et al (eds) Dredging, remediation and containment of contaminated sediments. ASTM Publication Code Number (PCN) 04-012930-38, Philadelphia

History Channel (2005) Modern marvels: dredging

Huston J (1970) Hydraulic dredging. Cornell Maritime Press, Cambridge

Iwasaki M, Kuioka K, Izumi S, Miyata N (1992) High density dredging and pneumatic conveying system. In: Proceedings of XIIIth world dredging congress, WODCON XIII, Bombay, 7–10 Apr 1992, pp 773–792

Jaglal K, McLaughlin DB (1999) Evaluation of large scale environmental dredging as a remedial alternative. In: Proceedings of the Western Dredging Association 19th technical conference and 31st Texas A & M Dredging seminar, Louisville, KY, 15–18 May 1999

Kikeqawa K (1983) Sand overlying for bottom sediment improvement by sand spreader. In: Management of bottom sediments containing toxic substances: proceedings of the 7th US/Japan Experts meeting. Prepared for the U.S. Army Corps of Engineers Water Resources Support Center by the U.S. Army Engineer Waterways Experiment Station, Vicksburg, pp 79–103

Landin MC (2000) Beneficial uses of dredged material. In: Herbich JB (ed) Handbook of dredging engineering. McGraw-Hill, New York, Chapter 16

LaSalle MW, Clarke DG, Homziak J, Lunz JD, Fredette TJ (1991) A framework for assessing the need for seasonal restrictions on dredging and disposal operations, Technical report D-91-1. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Lee CR (1997) Manufactured soil from toledo harbor dredged material and organic waste materials. In: International workshop on dredged material beneficial uses, Baltimore, MD, 28 July–1 Aug 1997

Marine Board (2001) A process for setting, managing and monitoring environmental windows for dredging projects. Transportation Research Board and Ocean Studies Board, National Research Council, National Academy Press, Washington, DC

Marine Board (1990) Managing troubled waters – The role of marine environmental monitoring. National Research Council, National Academy Press, Washington, DC

Marine Board, National Research Council (1997) Contaminated sediments in ports and waterways – cleanup strategies and technologies. National Academy Press, Washington, DC

Marine Board, National Research Council (1990) Managing coastal erosion. National Academy Press, Washington, DC

Marine Board, National Research Council (1987) Sedimentation control to reduce maintenance dredging of navigation facilities in Esturaries. National Academy Press, Washington, DC

Martin JL, McCutcheon SC (1992) Overview of processes affecting contaminant release from confined disposal facilities, Contract report D-92-1. U.S. Army Engineer Waterways Experiment Station, CE, Vicksburg, NTIS no AD A247 650

Matousek V (1997) Flow mechanism of sand-water mixtures in pipelines. Ph.D. Dissertation, Delft University of Technology, Delft, The Netherlands

McLellan TN, Havis RN, Hayes DF, Raymond GL (1989) Field studies of sediment resuspension characteristics of selected dredges, Technical report HL-89-9. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Miertschin M, Randall RE (1998) A general cost estimation program for cutter suction dredges. In: Poceedings of 15th world dredging congress, Las Vegas, NV, pp 1099–1115

Myers TE, Brannon JM, Tardy BA (1996) Leachate testing and evaluation for estuarine sediments, Technical report D-96-1. U.S. Army Engineer Waterways Experiment Station, Vicksburg, NTIS no AD A306 421

National Dredging Team (2003) Dredged material management: action agenda for the next decade, Jacksonville, FL, 23–25 Jan 2003

Ocean Studies Board, National Research Council, Transportation Research Board (2001) A process for setting, managing and monitoring environmental windows for dredging projects; Special report 262

Otis MJ, Andon S, Bellmer R (1990) New bedford harbor superfund pilot study: evaluation of dredging and dredged material disposal. US Army Engineer Division, New England

Palermo MR, Clausner JE, Rollings MP, Williams GL, Myers TE, Fredette TJ, Randall RE (1998) Guidance for subaqueous dredged material capping, Technical report DOER-1. U.S. Army Corps of Engineers, Washington, DC

Palermo MR, Maynard S, Miller J, Reible DD (1996) Guidance for in-situ subaqueous capping of contaminated sediments. Environmental Protection Agency, EPA 905-B96-004, Great Lakes National Program Office, Chicago, IL

Palermo MR, Schroeder PR, Estes TJ, Francingues NR (2008) Technical guidelines for environmental dredging of contaminated sediments, ERDC/EL TR-08-29. U.S. Army Engineer Research and Development Center, Vicksburg

Parchure TM, Sturdivant CN (1997) Development of a portable innovative contaminated sediment dredge, Final report, CPAR-CHL-97-2, Construction productivity advancement research (CPAR) program. USAE Waterways Experiment Station, Vicksburg

Parchure TM (1996) Equipment for contaminated sediment dredging, Technical report HL-96-17. USAE Waterways Experiment Station, Vicksburg

Pearson TH (1987) Benthic ecology in an accumulation sludge-disposal site. In: Capuzzo JM, Kester DR (eds) Ocean processes and marine pollution, vol 1, Biological processes and wastes in the ocean. Krieger Publishing, Malabar, pp 195–200

Pelletier JP (1992) Collingwood harbour sediment removal demonstration. Preliminary report on the water quality monitoring program, environmental protection, Environment Canada, Toronto, ON, Canada

Pennekamp JGS, Quaak MP (1990) Impact on the environment of turbidity caused by dredging, Terra et Aqua, 42. International Association of Dredging Companies, The Hague

Pequegnat WE, Gallaway BJ, Wright TD (1990) Revised procedural guide for designation surveys of ocean dredged material disposal sites, Technical report D-90-8. U.S. Army Engineer Waterways Experiment station, Vicksburg

PIANC (1992) Beneficial uses of dredged material: a practical guide. PIANC, Brussels/Belgium

Pianc, EnviCom (2009) Dredging management practices for the environment, Pianc report 100

Pilarczyk KW (1996) Geotextile systems in coastal engineering, an overview. In: Proceedings of the twenty-fifth international conference on coastal engineering, vol 2, pp 2114–2127

Pilarczyk KW (1994) Novel systems in coastal engineering, geotextile systems and other methods, an overview. Rijkswaterstaat, Delft

Randall RE (1994) Dredging equipment alternatives for contaminated sediment removal. In: Conference on environmental dredging technology, East Brunswick, NJ, Sponsored by The Port Authority of New York/New Jersey, 21–22 Jun 1994

Randall RE (2000) Estimating dredging costs. In: Herbich JB (ed) Handbook of dredging engineering, 2nd edn. McGraw-Hill, New York, Appendix 9

Randall RE (2004) Dredging. In: Tsinker G (ed) Handbook of port engineering. McGraw-Hill, New York, Chapter 11

Randall RE (2000) Pipeline transport of dredged material. In: Herbich JB (ed) Handbook of dredging engineering, 2nd edn. McGraw-Hill, New York, Chapter 7, Section 2

Randall RE (1992) Equipment used for dredging contaminated sediments. In: 1992 international environmental dredging symposium proceedings, Erie County Environmental Education Institute, Buffalo, NY, 30 Sep–2 Oct 1992

Randall RE, Clausner JE, Johnson BH (1994) Modeling of Cap placement at the New York mud dump site. In: Dredging 94, Second international conference of dredging and dredged material placement, ASCE, Orlando, FL, 13–16 Nov 1994

Randall R, Edge B, Basilotto J, Cobb D, Graalum S, He Q, Miertschin M (2000) Texas gulf intracoastal waterway (GIWW): beneficial uses, estimating costs, disposal analysis alternatives, and separation techniques, Texas Transportation Institute, Project report no 1733-S, Texas A&M University System, College Station, Texas, Sep 2000

Reine KJ, Dickerson DD, Clarke DG (1998) Environmental windows associated with dredging operations, DOER technical notes collection (TNDOER-E2). U.S. Army Engineer Research and Development Center, Vicksburg

Richardson M (2001) The dynamics of dredging. Placer Management Corp, Irvine, California

Sanders L, Killgore J (1989) Seasonal restrictions on dredging operations in freshwater systems, environmental effects of dredging. Technical note EEDP-01-16. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Sanderson WH, McKnight AL (1986) Survey of equipment and construction techniques for capping dredged material, Miscellaneous paper D-86-6. U.S. Army Engineer Waterways Experiment Station, Vicksburg

Santiago R (2000) Contaminated sediment removal general management options and support technologies, Dredging engineering short course notes, Center for Dredging Studies, Ocean Engineering Program, Civil Engineering Department, Texas A&M University, College Station, TX, Jan 2000

Schiller RE Jr (1992) Sediment transport in pipes. In: Herbich JB (ed) Handbook of dredging engineering. McGraw-Hill, New York, Chapter 6

Science Applications International Corporation (SAIC) (1995) Sediment capping of subaqueous dredged material disposal mounds: an overview of the New England experience, 1979–1993, DAMOS contribution #95, SAIC report no SAIC-90/7573&c84 prepared for the U.S. Army Engineer Division, New England, Waltham.

Shook CA, Rocco MC (1991) Slurry flow: principles and practice. Butterworth-Heinemann, Boston

Steevens J, Kennedy A, Farrar D, McNemar C, Reiss MR, Kropp RK, Doi J, Bridges T (2008) Dredged Material analysis tools; performance of acute and chronic sediment toxicity methods, ERDC/EL TR-08-16. U.S. Army Engineer Research and Development Center, Vicksburg

Sumeri A (1989) Confined aquatic disposal and capping of contaminated bottom sediments in puget sound. In: Proceedings of the WODCON XII, Dredging: technology, environmental, mining, world dredging congress, Orlando, FL, 2–5 May 1989

Togashi H (1983) Sand overlaying for sea bottom sediment improvement by conveyor barge. In: Management of bottom sediments containing toxic substances: proceedings of the 7th U.S./Japan experts meeting. Prepared for the U.S. Army Corps of Engineers Water Resources Support Center by the U.S. Army Engineer Waterways Experiment Station, Vicksburg, pp 59–78

Tsinker GP (2004) Port engineering: planning, construction, maintenance, and security. Wiley, Hoboken

Turner TM (1996) Fundamentals of hydraulic dredging, 2nd edn. ASCE Press, New York

USACE (1987a) Confined disposal of dredged material, Engineer manual, US Army Corps of Engineers, EM-1110-2-5027. US Government Printing Office, Washington, DC
USACE (1987b) Beneficial uses of dredged material, Engineer manual, US Army Corps of Engineers, EM-1110-2-5026. US Government Printing Office, Washington, DC
USACE (1983) Dredging and dredged material disposal, Engineer manual, US Army Corps of Engineers, EM-1110-2-5025. US Government Printing Office, Washington, DC
USEPA/USACE (1992) Evaluating environmental effects of dredged material management alternatives-A technical framework, EPA 842-B-92-008. US Government Printing Office, Washington, DC
USEPA/USACE (1996) Guidance document for development of site management plans for ocean dredged material disposal sites, USEPA, Office of Water, Washington, DC
USEPA/USACE (2007) The role of the federal standard in the beneficial use of dredged material from US Army Corps of Engineers new and maintenance navigation projects – Beneficial uses of dredged materials. USEPA/USACE, Washington, DC
USEPA/USACE (1996) Evaluation of dredged material proposed for discharge in inland and near-coastal waters – Testing manual. Office of Water, U.S. Environmental Protection Agency, Washington, DC
USEPA/USACE (1991) Evaluation of dredged material proposed for ocean disposal (Testing manual), EPA-503/8-91/001. Office of Water, U.S. Environmental Protection Agency, Washington, DC
Van Drimmelen C, Schut T (1992) New and adapted small dredges for remedial dredging operations. In: Proceedings of WODCON XIII, Bombay, India, pp 156–169
Vorhees DJ, Driscoll SBK, Stackelberg K, Bridges TS (1998) Improving dredged material management decisions with uncertainty analysis. U.S. Army Engineer Waterways Experiment Station, Vicksburg
Wilson KC, Addie GR, Sellgren A, Clift R (2006) Slurry transport using centrifugal pumps, 3rd edn. Blackie Academic, New York
Zappi PA, Hayes DF (1991) Innovative technologies for dredging contaminated sediments, Miscellaneous paper EL-91-20. U.S. Army Engineer Waterways Experiment Station, Vicksburg
Zeller RW, Wastler TA (1986) Tiered ocean disposal monitoring will minimize data requirements, Oceans '86, vol 3. Marine Technology Society, Washington, DC, pp 1004–1009

Driver Assistance System, Biologically Inspired

Moustapha Doumiati[1], Daniel Lechner[2]
[1]B2i Automotive Engineering Company, Massy, France
[2]Department of Accident Mechanism Analysis, IFSTTAR-MA Laboratory, Salon de Provence, France

Article Outline

Glossary
Definition of the Subject
Introduction
Estimation Process Description
Four-Wheel Vehicle Model
Tire–Road Interface
Observer's Design
Experimental Results
Conclusion and Future Perspectives
Acknowledgments
Bibliography

Glossary

Lateral tire forces They are responsible to hold on the vehicle during a turn.

Longitudinal tire forces They are responsible to accelerate/brake the vehicle.

Observer or estimator It models a real system in order to provide an estimate of its internal state, given measurements of the input and output of the real system.

Sideslip angle It is the angle between the velocity heading and the true heading of the vehicle.

Tire forces The developed forces (longitudinal and lateral) are function of tire properties (material, tread pattern, tread depth, profile, etc.), the normal load on the tire, and the velocities experienced by the tire.

Vehicle control systems They provide commands and instructions to control the movements of the vehicle in order to maintain stability and enhance passengers security and comfort.

Vehicle dynamics It includes analytical and experimental technology used to study and understand the dynamical responses of a vehicle in various in-motion situations.

Vertical (normal) tire forces They are responsible to support the weight of the vehicle.

Definition of the Subject

The principal concerns in driving safety with standard vehicles are understanding and preventing risky situations. A close examination of accident data reveals that losing the vehicle control is the main reason for most car accidents. To help the driver to prevent such accidents, vehicle control systems may be used. For their optimal operation, these control systems require certain input data concerning vehicle dynamic parameters

and vehicle–road interaction. Unfortunately, some fundamental parameters like the tire-road forces and the sideslip angle are difficult to measure in a car, for both technical and economic reasons. To face this problem, this entry presents a dynamic modeling and observation method to estimate these variables. The ability to accurately estimate lateral tire forces and sideslip angle is a critical determinant in the performances of many vehicle control systems. To address nonlinearities and unmodeled vehicle dynamics, an observer derived from unscented Kalman filtering technique is proposed. The estimation process method is based on the dynamic response of a vehicle instrumented with easily available and potentially integrable sensors. Performances are tested using an experimental car in real driving situations. Experimental results show the potential of the proposed estimation method.

Introduction

Vehicle dynamics and stability have been of considerable interest to automotive engineers, automobile manufacturers, government, public safety groups, and general public for a number of years. The obvious dilemma is that people naturally desire to drive faster and faster on the roads and highways, yet they expect their vehicles to be stable and safe during all normal and emergency maneuvers. For the most part, people pay little attention to the limited handling potential of their vehicles until some unusual behavior is observed that often results in fatality. According to statistics, worldwide, an estimated 1.2 million people are killed in road crashes each year and as many as 50 million are injured [1]. Preventing car accidents requires to know what determines vehicle dynamics during motion [2].

Today, automotive electronic technologies are developing for safe and comfortable traveling of drivers and passengers. Nowadays, there are a lot of vehicle control system such as the Anti-lock Braking System (ABS) that prevents wheel lock during braking [3], and the Electronic Stability Control (ESC) that enhances lateral vehicle stability [4, 5]. These control systems installation rate is increasing all around the world. Vehicle control algorithms have made great strides toward improving the handling and safety of vehicles. For example, experts estimate that ESC prevents 27% of loss-of-control accidents by intervening when emergency situations are detected [6]. While nowadays vehicle control algorithms are undoubtedly a life-saving technology, they are limited by the available vehicle state information.

Vehicle control systems currently available on production cars rely on available inexpensive measurements such as longitudinal velocity, accelerations, and the vehicle yaw rate. Sideslip rate can be evaluated using the yaw rate, lateral acceleration, and vehicle velocity [7]. Calculating the sideslip angle is possible from the sideslip rate integration. However, it is prone to uncertainty and errors from sensor bias. Besides, these control systems use unsophisticated, inaccurate tire models to evaluate lateral tire dynamics. In fact, measuring tire forces and sideslip angles is very difficult for technical and economic reasons. Therefore, these important data must be observed or estimated. If control systems were in possession of the complete set of lateral tire characteristics, namely, lateral forces, sideslip angle, and the tire–road friction coefficient, they could greatly enhance vehicle handling and increase passenger safety.

As the motion of a vehicle is governed by the forces generated between the tires and the road, knowledge of the tire forces is crucial when predicting vehicle motion. For example, a vehicle can turn because of the applied lateral tire forces. In fact, what happens is that when the front wheels of a vehicle are steered, a slip angle is created, which gives rise to a lateral force. This lateral force turns or yaws the vehicle. Under normal driving situations (low slip angle), a vehicle responds predictably to the driver's inputs. As the vehicle approaches the handling limits, for example, during an evasive emergency maneuver, or when a vehicle undergoes high accelerations, high slip angle occurs and the vehicle's dynamic becomes highly nonlinear and its response becomes less predictable and potentially very dangerous.

Accurate data about tire forces lead to a better evaluation of the vehicle possible trajectories, to a better vehicle control and rollover prevention. Moreover, it enables the development of a diagnostic tool for evaluating the potential risks of accidents related to poor adherence or dangerous maneuvers.

In the literature, many studies have looked at the vehicle dynamic states estimation. Several ones have

been conducted regarding the estimation of tire-road forces [8–20]. For example, in [8], a study of a 14-degrees-of-freedom (DOF) vehicle model is proposed where the dynamics of the roll center are used to calculate vertical tire forces. In [9], the authors propose an estimation method in order to estimate vertical and lateral forces per axle. The authors in [10–13] estimate the vehicle vertical forces and other dynamic states for a four-wheel vehicle model (FWVM) comprising four DOF. Consequently, lateral tire forces at each tire are calculated based on the estimated states and using a quasi-static tire model. In [14], Ray estimates the vehicle dynamic states and lateral tire forces per axle for a nine-DOF vehicle model. The author uses measurements of the applied torques as inputs to his model. We note that the torque is difficult to get in practice; it requires expensive sensors. More recently, authors in [15, 16] propose observers to estimate lateral forces per axle without using torque measures. In [17], the authors propose an estimation process based on a three-DOF vehicle model, as a lateral tire force estimator. In [15–17], lateral forces are modeled with a derivative equal to random noise. The authors in [17] remark that such modeling leads to a noticeable inaccuracy when estimating individual lateral tire forces, but not in axle lateral forces. This phenomenon is due to the non-representation of the lateral load transfer when modeling. Studies in [21–24] focus on the tire-road friction estimation.

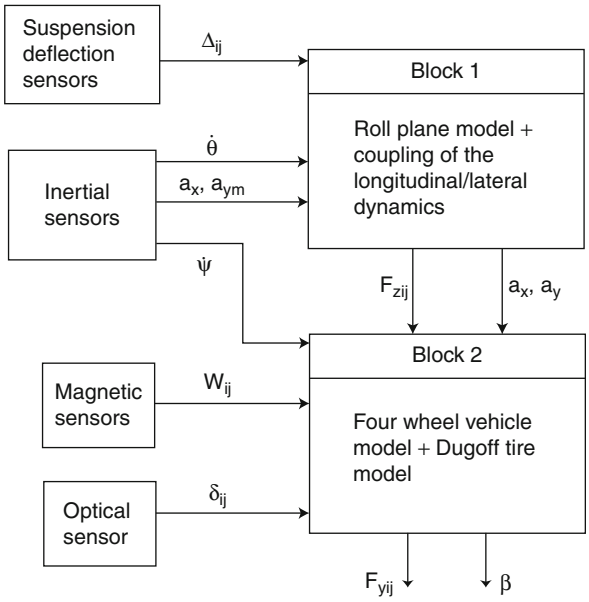
The main goal of this entry is to present an estimation method that uses simple vehicle-road models and a certain number of valid measurements in order to estimate in real time and in accurate way the sideslip angle and the lateral force at each individual tire–road contact point. This entry presents two significant particularities:

1. First, the estimation process does not use the measurements of wheel torques which are very expensive.
2. Second, the estimation process uses accurate normal tire forces, in contrast to many existing approaches that assume constant vertical forces. This approach is more realistic since during cornering, accelerating, and braking, the load distribution varies significantly in a car, thus cornering stiffness and lateral forces evaluation are directly affected.

The developed estimation process is model based and built using Kalman filter technique. The Kalman filter is known as the most commonly used real-time estimator for linear and nonlinear systems. In order to show the effectiveness of the estimation method, some real-time validation tests were carried out on an instrumented vehicle in realistic driving situations.

Estimation Process Description

The estimation process is shown in its entirety by the block diagram in Fig. 1, where a_x and a_{ym} are respectively the longitudinal and lateral accelerations, $\dot{\psi}$ is the yaw rate, $\dot{\theta}$ is the roll rate, Δ_{ij} (i represents the front 1 or the rear 2 and j represents the left 1 or the right 2) is the suspension deflection, w_{ij} is the wheel velocity, F_{zij} and F_{yij} are respectively the normal and lateral tire-road forces, β is the sideslip angle at the center of gravity (cog). The whole estimation process consists of two blocks, and its role is to estimate sideslip angle at the cog, normal and lateral forces at each tire–road contact point, and consequently evaluate the used lateral friction coefficient. The following measurements are needed:



Driver Assistance System, Biologically Inspired. Figure 1 Process estimation diagram

- Yaw and roll rates measured by gyrometers
- Longitudinal and lateral accelerations measured by accelerometers
- Suspension deflections using suspension deflections sensors
- Steering angle measured by an optical sensor
- Rotational velocity for each wheel given by magnetic sensors

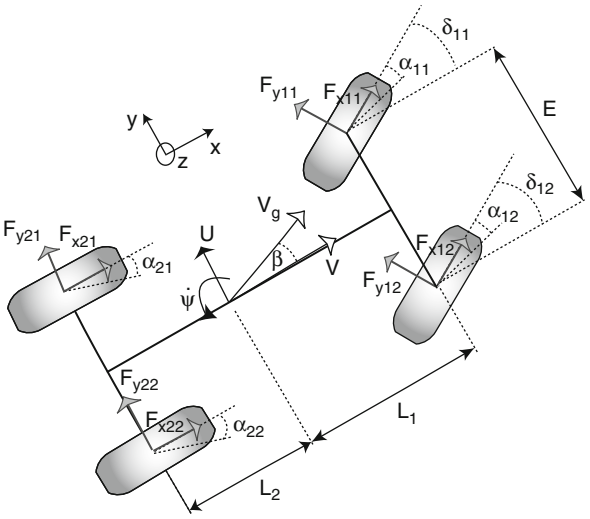
The first block aims to provide the vehicle's mass, lateral load transfer, normal tire forces, and the corrected lateral acceleration a_y (by canceling the gravitational acceleration component that distorts the accelerometer signal a_{ym}). It contains observers based on vehicle's roll dynamics and model that couples longitudinal and lateral accelerations. Authors have looked at the first block in some previous studies [10, 11]. This work focuses only on the second block, whose main role is to estimate individual lateral tire force and sideslip angle. The second block makes use of the estimations provided by the first block. In fact, as will be shown in the sections, the impact of including accurate normal forces in the calculation of lateral forces is fundamental.

One specificity of this estimation process is the use of blocks in series. By using cascaded observers, the observability problems entailed by an inappropriate use of the complete modeling equations are avoided, enabling the estimation process to be carried out in a simple and practical way. In the following, the model-based observer of the second block will be explained in details. Therefore, the vehicle-road model and the estimation method will be illustrated, respectively.

Four-Wheel Vehicle Model

The Four-Wheel Vehicle model (FWVM) is chosen for this entry because it is simple and corresponds sufficiently to our objectives. The FWVM is widely used to describe transversal vehicle dynamic behavior [12–14, 17, 18].

Figure 2 shows a simple diagram of the FWVM model in the longitudinal and lateral planes. In order to simplify the lateral and longitudinal dynamics, rolling resistance is neglected. Additionally, the front and rear track widths (E) are assumed to be equal. L_1 and L_2 represent the distance from the vehicle's center of gravity to the front and rear axles respectively. The



Driver Assistance System, Biologically Inspired. Figure 2 Four-wheel vehicle model

sideslip at the vehicle center of gravity (β) is the difference between the velocity heading (V_g) and the true heading of the vehicle (ψ). The yaw rate ($\dot{\psi}$) is the angular velocity of the vehicle about the center of gravity. The forward and lateral velocities are respectively $V$and U. The longitudinal and lateral forces ($F_{x,y,i,j}$) are shown for front and rear tires of the vehicle.

Longitudinal forces should be taken into account to enable accurate lateral forces estimation during vehicle braking or acceleration. While considering their effect is certainly important, its inclusion makes solving the lateral estimation problem considerably more complex. Thus, it may be desirable to solve the lateral estimation problem in the absence of longitudinal forces first and include them in later studies. This can be done by focusing on solving the estimation problem when the vehicle is driven at constant speeds [20].

This entry extended the hypothesis of moving in a constant speed and addresses the case of a front-wheel drive, where rear longitudinal forces are neglected relative to the front longitudinal forces. Longitudinal front axle forces are considered by assuming that:

$$F_{x1} = F_{x11} + F_{x12}. \tag{1}$$

The longitudinal force evolution is modeled with a random walk model, where its derivative is equal to random noise $\dot{F}_{x1} = 0$. This is due to the lack of

knowledge on the longitudinal slip and the effective radius of the tire [25].

The lateral dynamics of the vehicle can be obtained by summing the forces and moments about the vehicle's center of gravity. Consequently, the simplified FWVM is formulated as the following dynamic relationships [19]:

$$\begin{cases} \dot{V}_g = \frac{1}{m}\begin{bmatrix} F_{x1}\cos(\beta-\delta) + F_{y11}\sin(\beta-\delta)+ \\ F_{y12}\sin(\beta-\delta) + (F_{y21}+F_{y22})\sin\beta \end{bmatrix}, \\ \ddot{\psi} = \frac{1}{I_z}\begin{bmatrix} L_1[F_{y11}\cos\delta + F_{y12}\cos\delta + F_{x1}\sin\delta]- \\ L_2[F_{y21}+F_{y22}]+ \\ \frac{E}{2}[F_{y11}\sin\delta - F_{y12}\sin\delta] \end{bmatrix}, \\ \dot{\beta} = \frac{1}{mV_g}\begin{bmatrix} -F_{x1}\sin(\beta-\delta) + F_{y11}\cos(\beta-\delta)+ \\ F_{y12}\cos(\beta-\delta) + (F_{y21}+F_{y22}) \ \cos\beta \end{bmatrix} - \dot{\psi}, \\ a_y = \frac{1}{m}[F_{y11}\cos\delta + F_{y12}\cos\delta + (F_{y21}+F_{y22}) + F_{x1}\sin\delta], \\ a_x = \frac{1}{m}[-F_{y11}\sin\delta - F_{y12}\sin\delta + F_{x1}\cos\delta], \end{cases} \tag{2}$$

where m is the vehicle mass, and I_z is the yaw moment of inertia.

The tire slip angle (α_{ij}) as shown in Fig. 2, is the difference between the tire's longitudinal axis and the tire's velocity vector. The tire velocity vector can be obtained from the vehicle's velocity (at the cog) and the yaw rate. Assuming that rear steering angles are approximately null, the direction or heading of the rear tires is the same as that of the vehicle. The heading of the front tires includes the steering angle (δ). The front steering angles are assumed to be equal ($\delta_{11} = \delta_{12} = \delta$). The forward velocity V, steering angle δ, yaw rate $\dot{\psi}$, and the vehicle body slip angle β are then used to calculate the tire slip angles α_{ij}, where:

$$\begin{cases} \alpha_{11} = \delta - \arctan\left[\frac{V\beta + L_1\dot{\psi}}{V - E\dot{\psi}/2}\right], \\ \alpha_{12} = \delta - \arctan\left[\frac{V\beta + L_1\dot{\psi}}{V + E\dot{\psi}/2}\right], \\ \alpha_{21} = -\arctan\left[\frac{V\beta - L_2\dot{\psi}}{V - E\dot{\psi}/2}\right], \\ \alpha_{22} = -\arctan\left[\frac{V\beta - L_2\dot{\psi}}{V + E\dot{\psi}/2}\right]. \end{cases} \tag{3}$$

Tire–Road Interface

As the motion of a vehicle is governed by the forces generated between the tires and the road, knowledge of the tire forces is crucial in order to predict the vehicle's motion. This section presents the tire–road interaction phenomenon, especially the lateral tire forces. Since the quality of the observer largely depends on the accuracy of the tire model, the underlying model must be precise. Taking real-time calculation requirement, the tire model should also be simple.

Dugoff Tire Model

Many different tire models, based on the physical nature of the tire and/or on empirical formulations deriving from experimental data, can be found in the literature. These models include the Burckhardt, Dugoff, and Pacejka models [19, 26, 27]. One of the most commonly used model is the Pacejka's "Magic Formula." It is an effective method for predicting real tire behavior. However, it requires a large number of tire-specific parameters that are usually unknown. Another commonly used model is the Dugoff tire model. It synthesizes all the tire property parameters into two constants Cx and Cy, referred to as the longitudinal and cornering stiffness of the tire. Dugoff's model is the one used in this entry. Neglecting longitudinal forces, the simplified nonlinear lateral tire forces are given by:

$$\overline{F_{yij}} = -Cy_{ij}\, tan\alpha_{ij}\, f(\lambda), \tag{4}$$

where Cy_{ij} is the cornering stiffness, α_{ij} is the slip angle, and $f(\lambda)$ is given by:

$$f(\lambda) = \begin{cases} (2-\lambda)\lambda, & \text{if } \lambda < 1 \\ 1, & \text{if } \lambda < 1 \end{cases} \tag{5}$$

$$\lambda = \frac{\mu F_{zij}}{2C_{yij}|tan\alpha_{ij}|}. \tag{6}$$

In the above formulation, μ is the friction coefficient and F_{zij} is the vertical load on the tire. This simplified tire model assumes pure slip conditions with negligible longitudinal slip, a uniform pressure distribution, a rigid tire carcass, and a constant friction coefficient for sliding rubber. The original Dugoff tire model has

a constant stiffness in respect to weight transfer. It is worthy to note that, according to [28], load transfer affects the cornering stiffness Cy_{ij}. It can be represented by a second order polynomial with respect to the normal force as shown in equation (7):

$$Cy_{ij}(F_z) = \left(aFz_{ij} - bFz_{ij}^2\right), \tag{7}$$

where a and b are respectively the first and second order coefficient in the cornering stiffness polynomial. They are identified once using a set of experimental data treated offline, where the tires remain in their linear operation zone. Therefore, equation (7) is compared to the calculated cornering stiffness obtained from the ratio of the measured lateral force to the measured tire's slip angle. Hence, a and b are calculated, in such a way that the equation (7) fits the calculated cornering stiffness well.

This entry proposes a modified Dugoff tire model, where the cornering stiffness varies with respect to load.

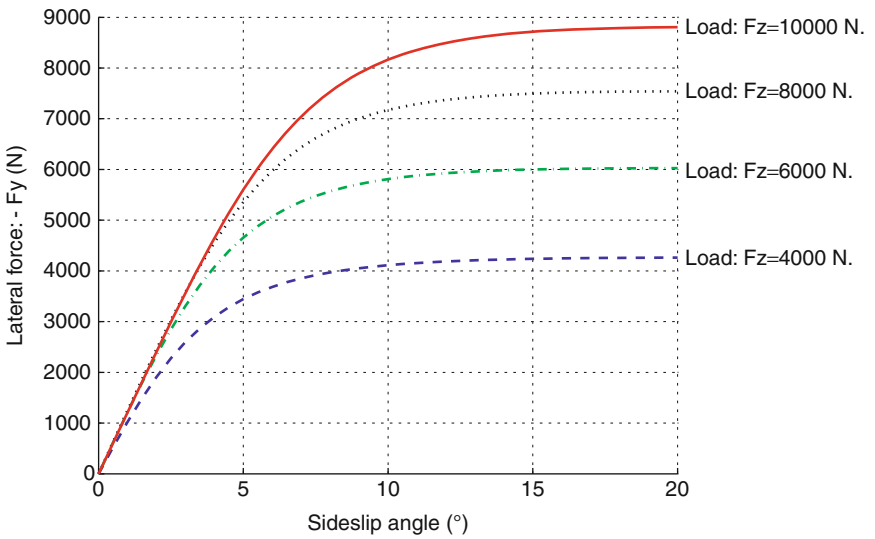
As shown in equation (4), vertical forces and the tire slip angles can be used to find the lateral force on each tire. Figure 3, based on the modified Dugoff's tire model, is a graph of the lateral force versus tire slip angle. It will be noted that as the load increases, the peak lateral force occurs at somehow higher slip angle.

It is clear that for small slip angles the force profile can be defined by a linear region. When operating in this region, a vehicle responds predictably to the driver's inputs.

As the slip angle continues to grow, the tire begins to saturate and reaches a peak value; this area is commonly called the nonlinear region of the tire curve. It represents the tire limits and it is rarely reached under normal driving conditions. If the front tires saturate first, the vehicle is said to display understeer, and may plow out of a bend. If the rear tires saturate first, the vehicle limit oversteers and may spin out. Because most drivers are not accustomed to operate in the nonlinear handling regime, both of these responses are potentially very dangerous. Noting that oversteer situation is much more difficult to be controlled than the understeer [29].

Observer's Design

This section presents a description of the observer devoted to lateral tire forces and sideslip angle. The state-space formulation, the observability analysis, and the estimation method will be presented.



Driver Assistance System, Biologically Inspired. Figure 3
Generic tire curve: lateral force versus slip angle

Stochastic State-Space Representation

The nonlinear stochastic state-space representation of the system described in previous section is given as:

$$\begin{cases} \dot{X}(t) = f(X(t),\ U(t)) + w(t) \\ Y(t) = h(X(t),\ U(t)) + v(t) \end{cases} \tag{8}$$

The input vector, U, comprises the steering angle and the normal forces considered estimated by the first block (see Fig. 1):

$$\begin{aligned} U &= [\delta,\ F_{z11},\ F_{z12},\ F_{z21},\ F_{z22}]^T \\ &= [u_1,\ u_2,\ u_3,\ u_4,\ u_5]^T. \end{aligned} \tag{9}$$

The measure vector, Y, comprises yaw rate, vehicle velocity (approximated by the mean of the rear wheel velocities calculated from wheel-encoder data), and longitudinal and lateral accelerations:

$$Y = [\dot{\psi},\ V_g,\ a_x,\ a_y]^T = [y_1,\ y_2,\ y_3,\ y_4]^T. \tag{10}$$

The state vector, X, comprises yaw rate, vehicle velocity, sideslip angle at the cog, lateral forces, and the sum of the front longitudinal tire forces:

$$\begin{aligned} X &= [\dot{\psi},\ V_g,\ \beta,\ F_{y11},\ F_{y12},\ F_{y21},\ F_{y22},\ F_{x1}]^T \\ &= [x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8]^T. \end{aligned} \tag{11}$$

Authors would like to emphasize that the consideration of the lateral forces as states allows:

1. A better evaluation of the tire forces. In fact, whatever the complexity of the tire models, there are several reasons why such models do not match the actual tire forces perfectly [30]. From these reasons, we can cite especially the changes in the tire's pressure and temperature and the changes in the road characteristics. Therefore, authors believe that according to the closed loop observer theory, the integration of the tire forces in the state vector may lead to better results than just using an open loop tire model.
2. A better understanding of the tire behaviors using the relaxation-length formulation, especially in transient maneuvers [31].
3. The forces reconstruction to be done robustly with respect to some parameter variations. In fact, it is well known that the Kalman filters have proven to be robust to parameter changes.

Taking these observations in mind, one can infer the contribution of this entry with respect to other existing studies in the literature like [12, 13], which estimate dynamic variables of the vehicle, and then assess the tire forces using a properly adjusted tire model.

The process and measurements noise vectors, respectively w and v, are assumed to be white, zero mean and uncorrelated.

Consequently, the particular nonlinear function $f(.)$ of the state equations is given by:

$$\begin{cases} f_1 = \dfrac{1}{I_z}\begin{bmatrix} L_1[x_4\ \cos\ u_1 + x_5\ \cos\ u_1 + x_8\ \sin\ u_1] - \\ L_2[x_6 + x_7] + \\ \dfrac{E}{2}[x_4\ \sin\ u_1 - x_5\ \sin\ u_1] \end{bmatrix}, \\ f_2 = \dfrac{1}{m}\begin{bmatrix} x_8\cos(x_3 - u_1) + x_4\sin(x_3 - u_1) + \\ x_5\sin(x_3 - u_1) + (x_6 + x_7)\ \sin(x_3) \end{bmatrix}, \\ f_3 = \dfrac{1}{mx_2}\begin{bmatrix} -x_8\sin(x_3 - u_1) + x_4\cos(x_3 - u_1) \\ +x_5\cos(x_3 - u_1) + (x_6 + x_7)\cos x_3 \end{bmatrix} - x_1, \\ f_4 = \dfrac{x_2}{\sigma_2}\left(-x_4 + \overline{F_{y11}}\ (\alpha_{11},\ u_2)\right), \\ f_5 = \dfrac{x_2}{\sigma_2}\left(-x_5 + \overline{F_{y12}}\ (\alpha_{12},\ u_3)\right), \\ f_6 = \dfrac{x_2}{\sigma_2}\left(-x_6 + \overline{F_{y21}}\ (\alpha_{21},\ u_4)\right), \\ f_7 = \dfrac{x_2}{\sigma_2}\left(-x_7 + \overline{F_{y22}}\ (\alpha_{22},\ u_5)\right), \\ f_8 = 0. \end{cases} \tag{12}$$

The observation function $h(.)$ is:

$$\begin{cases} h_1 = x_1, \\ h_2 = x_2, \\ h_3 = \dfrac{1}{m}[-x_4\ \sin\ u_1 - x_5\sin\ u_1 + x_8\ \cos\ u_1], \\ h_4 = \dfrac{1}{m}\ [x_4\cos\ u_1 + x_5\cos\ u_1 + (x_6 + x_7) \\ \qquad + x_6\ \sin\ u_1]. \end{cases} \tag{13}$$

The state vector $X(t)$ will be estimated by applying the unscented Kalman filter technique.

Observability

Observability is a measure of how well the internal states of a system can be inferred from knowledge of its inputs and external outputs. This property is often presented as a rank condition on the observability matrix. Using the nonlinear state-space formulation of the system presented in section "Stochastic State-Space Representation," the observability definition is local and uses the Lie derivative [32]. An observability analysis of this system was undertaken in [33]. It was shown that the system is observable except when:

- Steering angles are null
- The vehicle is at rest ($V_g = 0$)

For these situations, we assume that lateral forces and sideslip angle are null, which approximately corresponds to the real cases.

Estimation Method: Kalman Filter Algorithms

The aim of an observer or a virtual sensor is to estimate a particular unmeasurable variable from available measurements and a system model in a closed loop observation scheme, as illustrated in Fig. 4. Because of the vehicle system-model mismatches (unmodeled dynamics, parameter variations, etc.) and the presence of unknown and unmeasurable disturbances, the calculation obtained from vehicle's model would deviate from the actual values over time. In order to reduce the estimation error, at least some of the measured outputs are compared to the same variables estimated by the observer. The difference is fed back into the observer after being multiplied by a gain matrix K, and so we have a closed loop observer (see Fig. 4). All developed observers are implemented in a first-order Euler approximation discrete form. At each iteration, the state vector is first calculated according to the evolution equation and then corrected online with the measurement errors (innovation) and filter gain K in a recursive prediction-correction mechanism. In this study, the gain is calculated using the Kalman filter method which is a set of mathematical equations and is widely represented in [34–36].

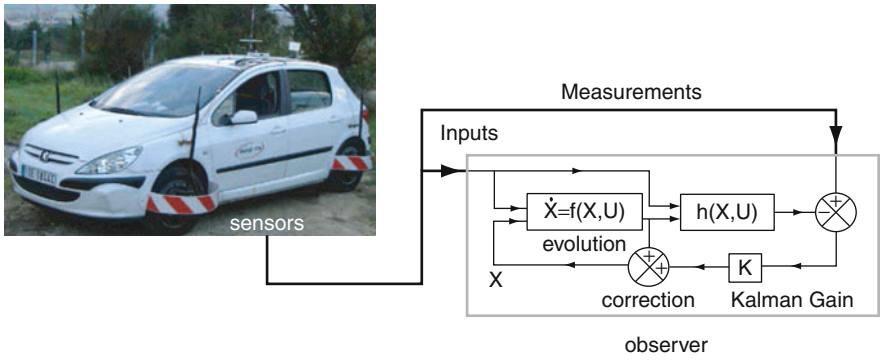
The EKF (Extended Kalman Filter) is probably the most commonly used estimator for nonlinear systems. However, in this study the UKF (Unscented Kalman Filter) [38–40] is chosen for the following fundamental reasons:

- The high nonlinearities of the vehicle-road model
- The calculation complexity of the Jacobian matrices which causes implementation difficulties

UKF Algorithm In this subsection, the principle of the UKF is summarized. Consider the general discrete nonlinear system:

$$\begin{cases} X_{k+1} = f(X_k,\ U_k) + w_k \\ Y_k = h(X_k) + v_k \end{cases} \tag{14}$$

where $X_k \in R^n$ is the state vector, $U_k \in R^r$ is the known input vector, $Y_k \in R^m$ is the output vector at time k. w_k and v_k are, respectively, the disturbance and sensor-noise vector, which are assumed to be Gaussian white noise with zero mean and uncorrelated.



Driver Assistance System, Biologically Inspired. Figure 4
Process estimation diagram

The UKF can be formulated as follows [37–40]:
Initialization

$$\begin{cases} \bar{X}_0 = E[X_0] \\ P_0 = E[(X_0 - \bar{X}_0)(X_0 - \bar{X}_0)^T] \end{cases} \tag{15}$$

where $\bar{X}_0$ and P_0 are respectively the initial state and the initial covariance.

Sigma points calculation and time update

$$\begin{cases} \chi_{k-1} = \left[\bar{X}_{k-1},\ \bar{X}_{k-1} \pm \sqrt{(n+\lambda)P_{k-1}}\right] \\ \chi^*_{k|k-1} = f(\chi_{k-1},\ U_{k-1}) \\ \bar{x}_{k|k-1} = \sum_{i=0}^{2n} w_i^m\ \chi^*_{i,k|k-1} \\ P_{k|k-1} = \sum_{i=0}^{2n} w_i^c \left(\chi^*_{i,k|k-1} - \bar{X}_{k|k-1}\right) \\ \qquad \times \left(\gamma^*_{i,k|k-1} - \bar{X}_{k|k-1}\right)^T + Q \\ \chi_{k|k-1} = \left[\bar{X}_{k-1},\ \bar{X}_{k-1} \pm \sqrt{(n+\lambda)P_{k|k-1}}\right] \\ \gamma_{k|k-1} = h(\chi_{k|k-1}) \\ \bar{Y}_{k|k-1} = \sum_{i=0}^{2n} w_i^m \gamma_{i,k|k-1} \end{cases} \tag{16}$$

where

$$\begin{cases} w_0^m = \dfrac{\lambda}{n+\lambda} \\ w_0^c = \dfrac{\lambda}{n+\lambda} + (n - \alpha^2 + \beta) \\ w_i^m = w_i^c = \dfrac{1}{2(n+\lambda)} \quad i = 1, \ldots, 2n \\ \lambda = n(\alpha^2 - 1) \end{cases} \tag{17}$$

Measurement update

$$\begin{cases} P\bar{Y}_k\bar{Y}_k = \sum_{i=0}^{2n} w_i^c(\gamma_{i,k|k-1} - \bar{Y}_{k|k-1}) \\ \qquad \times (\gamma_{i,k|k-1} - \bar{Y}_{k|k-1})^T + R \\ P_{\bar{X}_k\bar{Y}_k} = \sum_{i=0}^{2n} w_i^c(\chi_{i,k|k-1} - \bar{X}_{k|k-1}) \\ \qquad \times (\gamma_{i,k|k-1} - \bar{Y}_{k|k-1})^T \\ K_k = P_{\bar{X}_k\bar{Y}_k} P^{-1}_{\bar{Y}_k\bar{Y}_k} \\ P_k = P_{k|k-1} - K_k P_{\bar{Y}_k\bar{Y}_k} K_k^T \\ \bar{X}_k = \bar{X}_{k|k-1} + K_k(Y_k - \bar{Y}_{k|k-1}) \end{cases} \tag{18}$$

where the variables are defined as follows: $\bar{X}_k$ and $\bar{Y}_{k|k-1}$ are the estimations respectively of the state and of the real measurement at each instant k. w_i is a set of scalar weights, and n is the state dimension; the parameter α determines the spread of the sigma points around $\bar{X}$ and is usually set to $1e-4 < \alpha < 1$. The constant β is used to incorporate part of the prior knowledge of the distribution of X, and for Gaussian distributions, $\beta = 2$ is optimal. Q and R are respectively the disturbance and sensor-noise covariance: R takes into account the uncertainty in the measured data and Q is tuned depending on the model quality. Remember that the computation of the Kalman gain is a subtle mix between process and observation noises. The less noise in the operation compared to the uncertainty in the model, the more the variables will be adapted to follow measurements. Since the lateral forces are modeled using a relaxation model based on reliable tire models, the uncertainty affected to them is not too high. However, the longitudinal force per front axle is not modeled at all; hence, it is represented by a high noise level. The other states (yaw rate, longitudinal and lateral vehicle velocity) are modeled using the vehicle's equations. Therefore, they are said to have an average noise. On the other hand, since the embedded sensors have good accuracy, the noises on the measurements are quite small. In order to reduce the complexity of the problem, both measurement covariance matrix, R, and the process covariance matrix, Q, are assumed to be constant and diagonal. The off-diagonal elements are set to 0. That means that both the measurement noises and the process noises are supposed uncorrelated.

Experimental Results

In this section, the experimental car used to evaluate the observer performances is presented. Moreover, the test conditions and the results of the previously developed observers are discussed and analyzed.

Experimental Car

The experimental vehicle shown in Fig. 4 is the INRETS-MA (Institut National de Recherche sur les Transports et leur Sécurité – Département Mécanismes d'Accidents) Laboratory's test vehicle [7]. It is a Peugeot 307 equipped with:

- Gyrometers and accelerometers that measure respectively the rotations (roll, pitch, and yaw

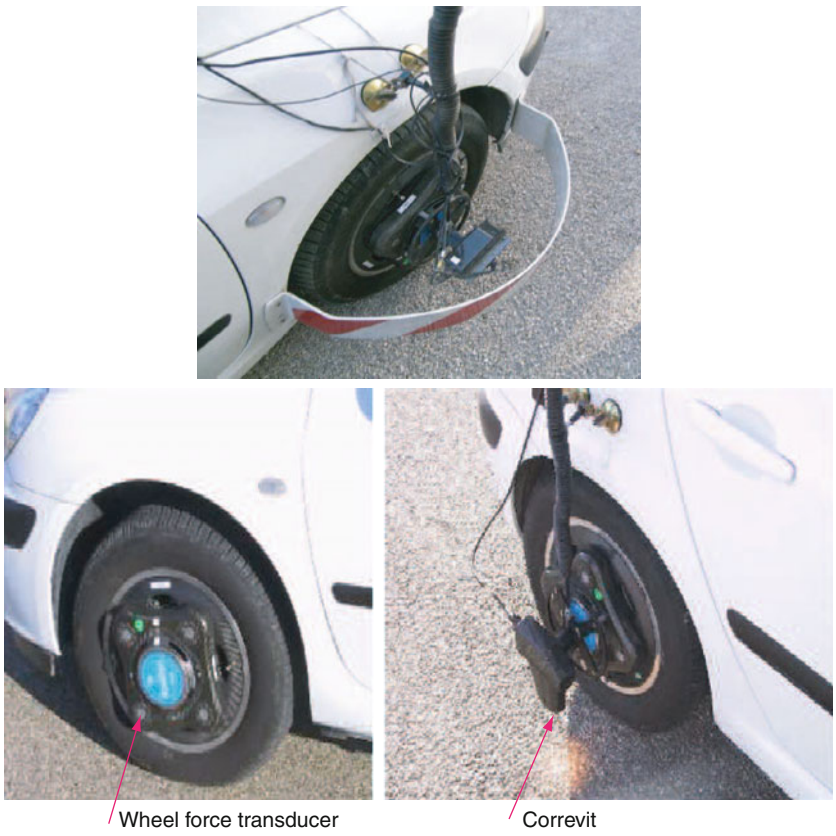
rates) and accelerations (longitudinal, lateral, and vertical) of the car body

- Suspension sensors that measure the distances between the wheels and the car body
- Three Correvit non-contact optical sensors:
  1. One is located in chassis rear overhanging position, and it measures longitudinal and lateral vehicle speeds.
  2. The other two are installed on the front right and rear right tires and they measure front and rear tires' longitudinal/lateral velocities and sideslip angles.
- Dynamometric wheels fitted on all four tires, which are able to measure tire forces and wheel torques in and around all three dimensions
- Steering-rack displacement sensor that is used to determine the steering angle
- Magnetic sensors that measure rotational velocity for each wheel

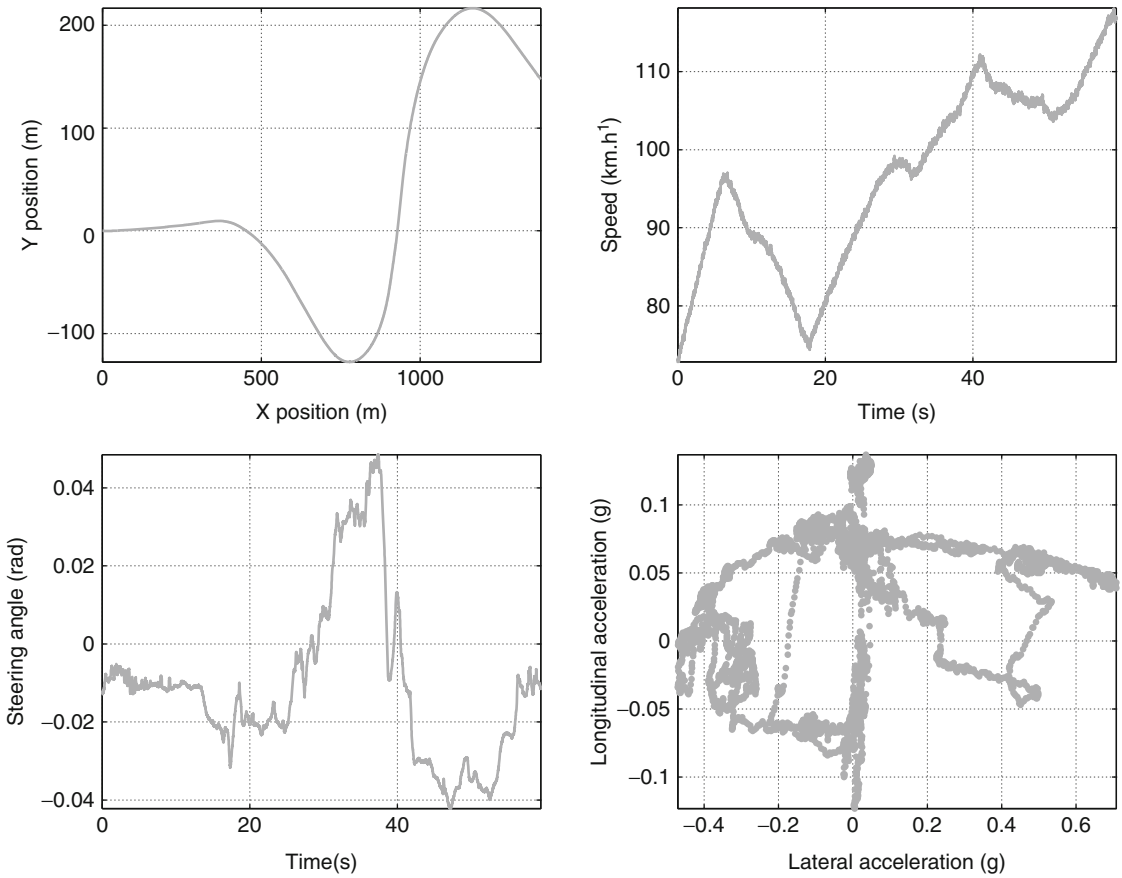
It is important to note that the Correvit and the wheel-force transducers (see Fig. 5) are very expensive sensors. They are used in this study as a reference for validating the estimation process. The sampling frequency of the different sensors is 100 Hz.

Test Conditions

Test data from nominal as well as adverse driving conditions were used to assess the performance of the observer presented in section "Observers Design," in realistic driving situations. We report a lane-change maneuver where the dynamic contributions play an important role. Figure 6 presents the Peugeot's trajectory (on a dry road), its speed, steering angle, and "g-g"



Driver Assistance System, Biologically Inspired. Figure 5
Wheel-force transducer and sideslip sensor installed at the tire level



Driver Assistance System, Biologically Inspired. Figure 6
Experimental test: vehicle trajectory, speed, steering angle, and acceleration diagrams for the lane-change test

acceleration diagram during the course of the test. The acceleration diagram, which determines the maneuvering area utilized by the driver/vehicle, shows that large lateral accelerations were obtained (absolute value up to 0.6 g). This means that the experimental vehicle was put in a critical driving situation.

The estimation process algorithm was written in C++ and has been integrated into the laboratory car as a DLL (Dynamic Link Library) that functions according to the software acquisition system.

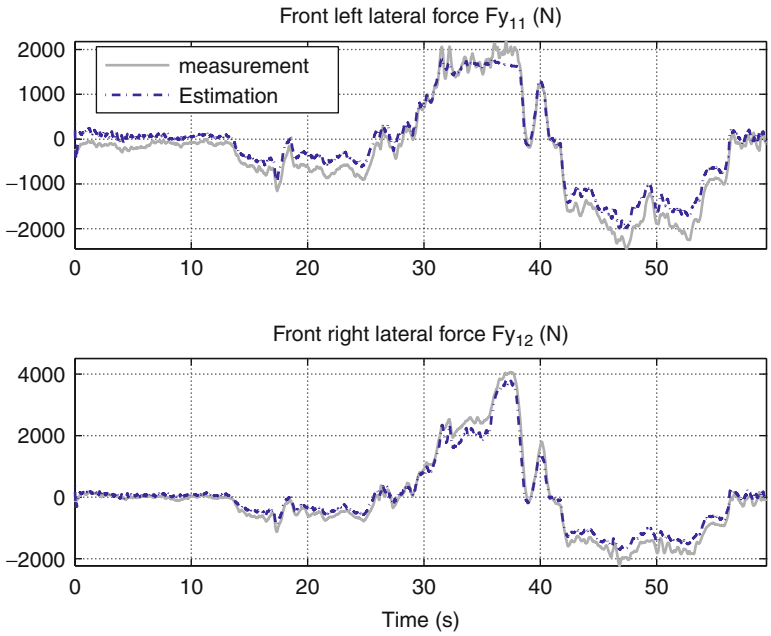
Validation of Observers

The observer results are presented in two forms: as tables of normalized errors and as figures comparing the measurements and the estimations. The normalized error for an estimation z is defined as:

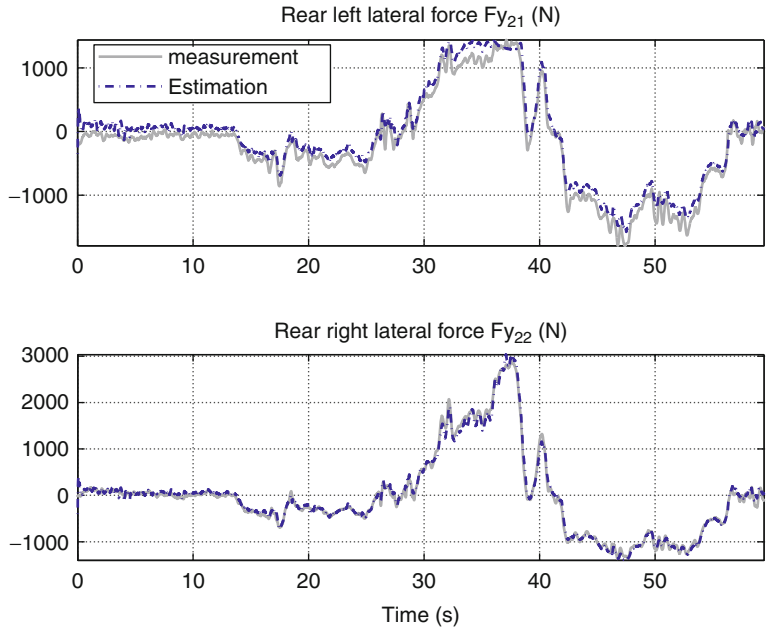
$$\epsilon_z = 100 \times \frac{\| z_{obs} - z_{measured} \|}{max(\| z_{measured} \|)} \quad (19)$$

where z_{obs} is the variable calculated by the observer, $z_{measured}$ is the measured variable, and $max(||z_{measured}||)$ is the absolute maximum value of the measured variable during the test maneuver.

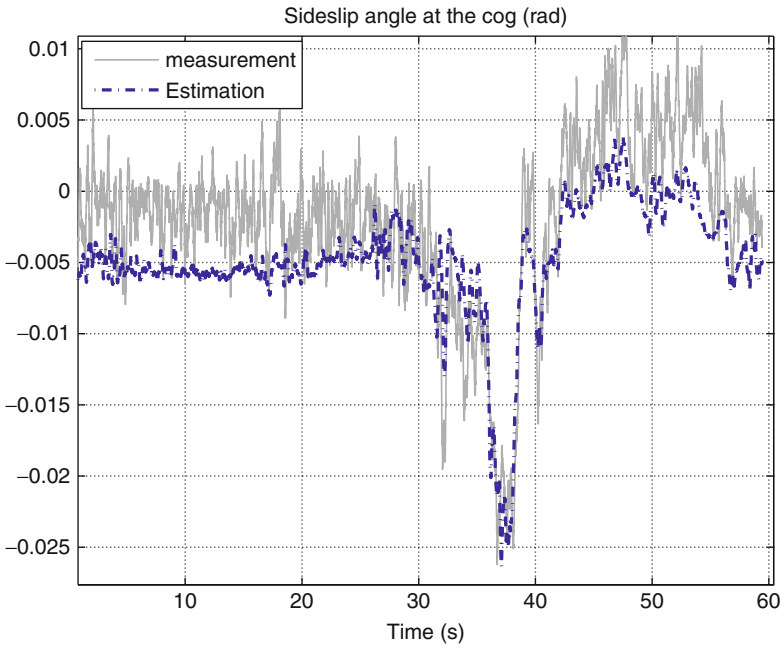
Figures 7 and 8 show lateral forces on the front and rear wheels. According to these plots, the observers are relatively good with respect to measurements. Some small differences during the trajectory are to be noted. These might be explained by neglected geometrical parameters, especially the camber angles, which also produce a lateral forces component [41]. It is also shown that the lateral forces on the right-hand tires exceed those on the left-hand tires. This result is clearly



Driver Assistance System, Biologically Inspired. Figure 7
Estimation of front lateral tire forces



Driver Assistance System, Biologically Inspired. Figure 8
Estimation of rear lateral tire forces



Driver Assistance System, Biologically Inspired. Figure 9
Estimation of the sideslip angle at the cog

a consequence of the load transfer produced during cornering from the left to the right-hand side of the vehicle. In fact, lateral force increases as normal force increases. Figure 9 shows how sideslip angle changes during the test. Reported results are relatively good.

Table 1 presents maximum absolute values, normalized mean errors, and normalized std (standard deviations) for lateral tire forces and sideslip angles. Despite the simplicity of the model, we can deduce that for this test, the performance of the observer is satisfactory, with normalized error globally less than 8%.

Given the vertical and lateral tire forces at each tire–road contact level, the estimation process is able to evaluate the used lateral friction coefficient μ. This is defined as the ratio of friction force to normal force, and is given by [29, 41]:

$$\mu_{ij} = \frac{Fy_{ij}}{Fz_{ij}} \tag{20}$$

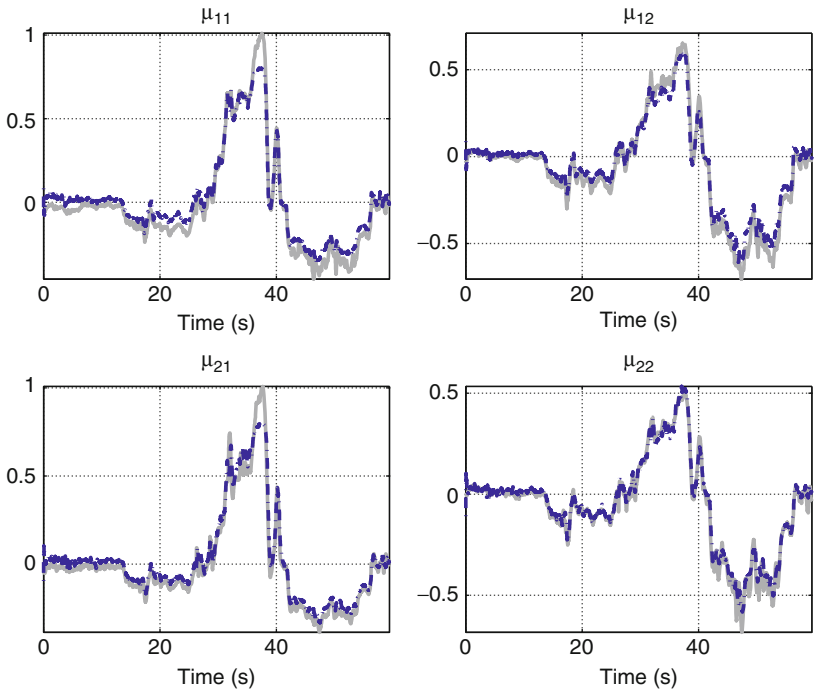
The lateral friction coefficients in Fig. 10 show that the estimated μ_{ij} are close to the measured values.

Driver Assistance System, Biologically Inspired. Table 1
Maximum absolute values, normalized mean errors, and normalized std

	Max || ||	Mean %	Std %
F_{y11}	2180 (*N*)	8.23	4.80
F_{y12}	4070 (*N*)	3.70	3.74
F_{y21}	1441 (*N*)	7.51	3.52
F_{y22}	2889 (*N*)	1.91	1.77
β	0.027 (*rad*)	8.32	6.41

A closer investigation reveals that the used lateral friction coefficients μ_{12} and μ_{22} corresponding to the overloaded tires during cornering are lower than μ_{11} and μ_{21}. This phenomenon is due to the tire load sensitivity effect: the lateral friction coefficient is normally higher for the lighter loads, or conversely, falls off as the load increases [29, 41].

This test also demonstrates that μ_{11} and μ_{21} are high, especially for lateral accelerations up to 0.6,



Driver Assistance System, Biologically Inspired. Figure 10
Used lateral friction coefficients developed by the tires

and that they attain the limit for the dry road friction coefficient. In fact, dry road surfaces show a high friction coefficient in the range 0.9–1.2 (implying that driving on these surfaces is safe), which means that for this test the limits of handling were reached.

Conclusion and Future Perspectives

This entry presents an interesting method for estimating lateral tire forces and sideslip angle, that is to say two of the most important parameters affecting vehicle stability and the risk of leaving the road. Consequently, the developed observer could feed control systems with fundamental vehicle-dynamics data in order to enhance vehicle safety.

The proposed observer is derived from a simplified four-wheel vehicle model and is based on unscented Kalman filtering technique. Tire–road interaction is represented by the Dugoff model. A comparison with real experimental data demonstrates the potential of the estimation process. It is shown that it may be possible to replace expensive correvit and dynamometric hub sensors by real-time software observers. This is one of the important results of this report. Another important result concerns the estimation of individual lateral forces acting on each tire. This can be seen as an advance with respect to the current vehicle-dynamics literature.

Future studies will improve the vehicle-road model in order to widen validity domains for the observer (take into account road irregularities and road bank angle), and make it adaptive with the road conditions (especially the road friction). Moreover, it will be of major importance to study the effect of coupling longitudinal/lateral dynamics on lateral tire behavior.

Acknowledgments

This work was done at Heudiasyc Laboratory UMR CNRS 6599, UTC University (Compiègne, France), in collaboration with Alessandro Victorino and Ali Charara.

Bibliography

Primary Literature

1. FARS (2007) Fatality analysis reporting system (FARS). Technical report, National Highway Traffic Safety Administration, National Center for Statistics and Analysis
2. Aparicio F, Paez P, Moreno F, Jimenez F, Lopez A (2005) Discussion of a new adaptive speed control system incorporating the geometric characteristics of the road International. J Veh Auton Syst 3(1):47–64
3. SAE Society of Automotive Engineers (2004) Bosch automotive handbook, 6th edn. SAE Society of Automotive Engineers, Warrendale
4. ESC (2003) Electronic stability control fact sheet. Technical report, ESC Coalition and DEKRA Automotive research
5. Van Zanten A (2002) Evolution of electronic control systems for improving the vehicle dynamic behavior. In: Proceedings of the 6th international symposium on advanced vehicle control (AVEC), Hiroshima, pp 7–15
6. Hsu Y, Gerdes JC (2005) Stabilization of a steerby-wire vehicle at the limits of handling using feedback linearization. In: Proceedings of IMECE 2005, AMSE international mechanical engineering congress and exposition, Florida
7. Lechner D (2008) Embedded laboratory for vehicle dynamic measurements. In: International symposium on advanced vehicle control, Kobe
8. Shim T, Ghike G (2007) Understanding the limitations of different vehicle models for roll dynamics studies. Veh Syst Dyn 45:191–216
9. Doumiati M, Baffet G, Lechner D, Victorino A, Charara A (2008) Embedded estimation of the tire/road forces and validation in a laboratory vehicle. International symposium on advanced vehicle control, Kobe
10. Doumiati M, Victorino A, Charara A, Lechner D, Baffet G (2008) An estimation process for vehicle wheelground contact normal forces. In: IFAC world congress, Seoul
11. Doumiati M, Victorino A, Charara A, Lechner D (2009) Lateral load transfer and normal forces estimation for vehicle safety experimental evaluation. Veh Syst Dyn 47(12):1511–1533
12. Wenzel TA, Burnham KJ, Blundell MV, Williams RA (2006) Estimation of the nonlinear suspension tyre cornering forces from experimental road test data. Veh Syst Dyn 44:153–171
13. Dakhlallah J, Glaser S, Mammar S, Sebsadji Y (2008) Tire-road forces estimation using extended Kalman filter and sideslip angle evaluation. American Control Conference, Seatle
14. Ray LR (1997) Nonlinear tire force estimation and road friction identification: simulation and experiments. Automatica 33(10):1819–1833
15. Baffet G, Charara A, Dherbomez G (2007) An observer of tire road forces and friction for active-security vehicle systems. Mechatronics, IEEE/ASME 12:651–661
16. Baffet G, Charara A, Lechner D, Thomas D (2008) Experimental evaluation of observers for tire-road forces, sideslip angle and wheel cornering stiffness. Veh Syst Dyn 45:191–216
17. Wilkin MA, Manning WJ, Crolla DA, Levesley MC (2006) Use of an extended Kalman filter as a robust tyre force estimator. Veh Syst Dyn 44:50–59
18. Venhovens PT, Naab K (1999) Vehicle dynamics estimation using Kalman filters. Veh Syst Dyn 32:171–184
19. Kiencke U, Daiss A (1997) Observation of lateral vehicle dynamics. Control Eng Pract 5(8):1145–1150
20. Hsu YHJ (2009) Estimation and control of lateral tire forces using steering torque. Ph.D., Department of mechanical engineering, Stanford University
21. Haffner L, Kozek M, Shi J, Jorgel HP (2008) Estimation of the maximum friction coefficient for a passenger vehicle using the instantaneous cornering stiffness. In: American control conference, Seatle
22. Gustafsson F (1997) Slip-based tire-road friction estimation. Automatica 33(6):1087–1099
23. Hahn JO, Rajamani R, Alexander L (2002) GPS-Based real time identification of tire-road friction coefficient. IEEE Trans Control Syst Technol 10(3):331–343
24. Uchanski MR (2001) Road friction estimation for automobiles using digital signal processing methods. Ph.D. thesis, University of California, Berkley
25. Koo SL, Tan HS (2007) Tire dynamic deflection and its impact on vehicle longitudinal dynamics and control. Mechatronics IEEE/ASME 12(6):623–631
26. Dugoff J, Fanches P, Segel L (1970) An analysis of tire properties and their influence on vehicle dynamic performance. Society of Automotive Engineers paper: 700377
27. Pacejka HB (2002) Tyre and vehicle dynamics. Elsevier, Amsterdam
28. Gillespie TD (1992) Fundamental of vehicle dynamics. SAE, Warrendale
29. Rajamani R (2005) Vehicle dynamics and control. Springer, New York
30. Lidner L (1993) Experience with the magic formula tyre model. In: Proceedings of the 1st international colloquium on tyre models for vehicle dynamic analysis, Amsterdam
31. Svendenius J (2007) Tire modeling and friction estimation. Ph.D. thesis, Lund University, Sweden
32. Nijmeijer H, Van der Schaft AJ (1991) Nonlinear dynamical control systems. Springer, New York
33. Doumiati M, Victorino A, Charara A, Lechner D (2009) Unscented Kalman filter for real-time vehicle lateral tire forces and sideslip angle estimation. In: IEEE intelligent vehicles symposium, Xi'an
34. Kalman RE (1960) A new approach to linear filtering and prediction problems. TransAS ME-J Basic Eng 82(D):35–45
35. Welch G, Bishop G (2001) An introduction to the Kalman filter. Course 8 University of North Carolina, Department of Computer Science
36. Durrant-Whyte H (2001) Multi sensor data fusion. Technical report, Australian center for field robotics, University of Sydney
37. Mohinder GS, Angus PA (1993) Kalman filtering theory and practice. Prentice Hall, Englewood Cliffs
38. Julier SJ, Uhlman JK (1997) A new extension of the Kalman filter to nonlinear systems. In: International symposium aerospace/defense sensing, simulation and controls, Orlando

39. Julier SJ, Uhlman JK, Durrant-Whyte HF (1995) A new extension approach for filtering nonlinear systems. In: American control conference, Seattle
40. Wan EA, Merwe RVD (2000) The unscented Kalman filter for nonlinear estimation. In: Proceedings of the adaptive systems for signal processing, communications, and control symposium, Atlanta, pp 153–158
41. Milliken W, Milliken DL (1995) Race car vehicle dynamics. SAE, Warrendale

Books and Reviews

Aga M, Okada A (2003) Analysis of vehicle stability control (VSC)'s effectiveness from accident data. In: Proceedings of the 18th international technical conference on the enhanced safety of vehicles, Nagoya

Aguirre LA, Letellier C (2005) Observability of multivariate differential embeddings. J Phys A Math Gen 38:6311–6326

Bakker E, PAcejka HB, Lidner L (1987) A new tire model with application in vehicle dynamic systems. SAE paper 890087

Besançon G (2007) Nonlinear observers and applications. Springer, Berlin

Bevly DM, Gerdes JC, Wilson C, Zheng G (2000) The use of GPS based velocity measurements for improved vehicle state estimation. In: Proceedings of the American control conference, Chicago

Brown T, Hac A, Martens J (2004) Detection of vehicle rollover. In: SAE world congress, Michigan

Doumiati M, Sename O, Martinez J, Poussot C, Dugard L (2010) Gain scheduled steering and braking control for improvement of vehicle handling and stability. In: IEEE conference on decision and control, Georgia

Doumiati M, Sename O, Martinez J, Dugard L, Gaspar P, Szabo Z, Bokor J (2011) Vehicle yaw control via coordinated use of steering/braking systems. In: IFAC world congress, Milano

Heydinger G, Howe J (2000) Analysis of vehicle response data measured during severe maneuvers. SAE paper 2000-01-1644

Jazar RN (2008) Vehicle dynamics, theory and application. Springer, New York

Smith ND (2004) Understanding parameters influencing tire modeling. Colorado State University, Formula SAE Platform

Driver Assistance Systems, Automatic Detection and Site Mapping

Andreas Wimmer, Klaus C. J. Dietmayer
Institute of Measurement, Control, and Microtechnology, University of Ulm, Ulm, Germany

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Sensors and Measurement Vehicle
Mapping of Safety Barriers
Detection of Beacons and Traffic Pylons
Detection and Mapping of Road Markings
Highly Accurate Road Work Map
Future Directions
Bibliography

Glossary

Advanced driver assistance system Driver assistance system with environmental perception which warns or informs the driver or even activates actuators for, for example, braking or steering.

Beacon, traffic pylon, guiding reflector post Infrastructure elements, which are often used at road construction sites for guiding motorists at areas where the traffic routing has changed or at potentially dangerous spots.

Environmental perception Concept of measuring and evaluating the environment, here: the surrounding of a vehicle.

Global Navigation Satellite System (GNSS) Satellite-based system for positioning in a global reference system, e.g., GPS, GLONASS.

Laser scanner Range sensor for environmental perception with a rotating laser beam and high angular resolution. The measurement principle is based on the time of flight of light.

Road construction site Area where a road is built or rehabilitated, also: road works, work zone.

Road marking Marking on the road to inform drivers, e.g., to delineate the traffic lanes.

Safety barrier Infrastructure element for physically separating traffic from an opposing lane or work zone, e.g., crash barrier/guard rail, Jersey barrier.

Definition of the Subject and Its Importance

Road construction sites are often the reason for traffic congestion and accidents on highways and freeways. This causes great economic and ecological costs to the society and environment through increasing travel time and additional fuel consumption. Driver assistance systems specifically designed for work zones will help to reduce the negative impact of construction sites for traffic flow. At road works, the lane width usually is

reduced, which makes lane keeping a challenging task, especially for heavy duty vehicles. It often happens that truck drivers slightly ride over the lane markings, thus preventing other vehicles to use the neighboring lane at dual carriageways. The aim is to provide lane keeping support for vehicles even in complex scenarios like road construction sites. An assistance system which laterally controls a heavy-duty vehicle highly depends on a robust and accurate estimation of the position of the vehicle within its lane. Depending on weather and lighting condition, this may not be achievable with a camera sensor only. Therefore, a system is proposed which additionally uses a laser scanner and a precise and detailed digital road work map, which might be regarded as an extension to standard navigation maps. This highly accurate map may be sent to vehicles via wireless communication technique as soon as they approach the road construction site. The following chapters describe how the layout and all typical elements of work zones on highways and freeways can be detected with modern sensors. Then, this data is used to automatically generate the above-mentioned highly accurate road work map.

Introduction

Mobility plays an important role in the society: people travel to work and goods are transported. The freight transport volume on roads constantly increases in many countries worldwide [1, 2]. On the one hand, the additional number of cars contributes to a higher possibility of congestion. On the other hand, heavy-duty vehicles cause high stress on the road surface, more than much lighter passenger cars. This makes it necessary to constantly maintain and repair the carriageways. On account of this, a road construction sites is set up.

Additionally, constructing new highways and freeways as well as the rehabilitation of existing ones is an accepted part of economic stimuli programs, as investments in the infrastructure are believed to have a high return on investment [3]. As a result, the number of road work zones is likely to increase in the future.

Road construction sites, on the contrary, account for additional stress and annoyance of car drivers due to the higher risk of traffic jams and delays caused by reduced speed limits. There is a survey, which reports an increased nervousness of drivers while traveling through a work zone [4]. Another survey on 1,700 car drivers [5] revealed that 41% regard a construction site as source of danger. Especially the reduced lane width is a reason why drivers feel uncomfortable.

If the number of lanes is reduced, the total throughput capacity may decrease as well, which is another reason for congestion. But not only discomfort of road users is an issue, there are several studies which identify a higher accident rate at a road work zones. Khattak et al. [6] extensively evaluated the effect of work zone presence on crashes in the USA. Approximately 24,000 injury crashes and 700 fatalities occur at US road construction sites every year. In this work, 36 observations have been examined, both on prework zone and during-work zone. They ascertained that the total work zone crash rate on highways is 21.5% higher than the prework zone crash rate. Similar analyses have been performed in other countries. The ARROWS project (Advanced Research on Road Work Zone Safety Standards in Europe [7]) states that "work zones generally have higher accident rates than the same road sections without a work zone". In detail, however, the data differs between countries [7, 8]. The main reasons for these accidents are driving errors and have been examined in [9]. Motorists have to recognize the work zone, assess geometry of their lane, and predict the behavior of other road users in dense traffic. An inadequate or late response may easily result in crashes [9].

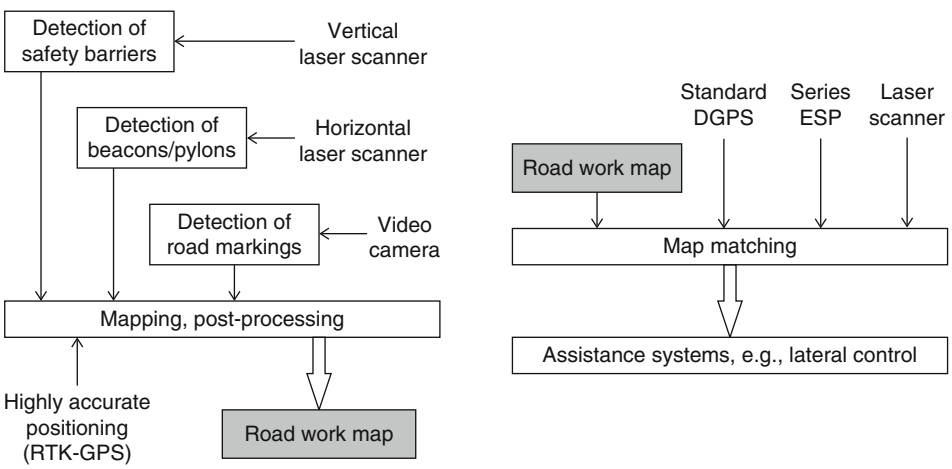
The question arises, what can be done to make work zones less dangerous and stressful for drivers. Accidents statistics [10] show that rear-end collisions represent a high relative portion. This issue is addressed by driver assistance systems such as ACC (Adaptive Cruise Control) or AEB (Automatic Emergency Break) [11, 12], which are either already in series production or about to be released. But at work zones, sideswipe collisions occur more often than on regular roads. An assistance system which helps centering the vehicle within its lane will reduce the possibility for accidents of this type. There are several systems available, such as Lane Keeping Assist (LKA) which are designed for standard highways and freeways. The layout of a road construction site usually is much more complex and thus poses a great challenge for driver assistance systems. The following contribution deals with the special needs to

automatically interpret work zones for future driver assistance systems. This concept should serve several purposes: aligning the vehicle within its lane reduces the work load for the driver and thus enabling him to concentrate on the surrounding traffic, which reduces the risk of accidents. If a heavy duty vehicle is optimally laterally controlled, there will be enough space on the neighboring lane so that other vehicles can easily pass. This again reduces the stress on drivers and additionally helps to maintain the full throughput capacity as all lanes can be used.

The above-mentioned goals rely on a robust and accurate estimation of the position of the vehicle within the work zone and especially within its lane. The required robustness can be achieved by fusing information of different sensors. A single video camera designed for lane detection may not be able to detect and interpret the complex scenery of a road construction site if the lighting condition and weather is disadvantageous. Rain and thus a wet surface may produce reflections on the road, which might be misinterpreted as road marking. Adverse lighting, such as direct sunlight or fog, often decreases the contrast of the image. Longitudinal groove zones might erroneously be interpreted as lane markings and therefore lead to false detections. In dense traffic, some road markings may also be not detectable due to occlusions from other vehicles. If the road surface is covered by snow, a video-based detection of the road markings is not possible. For these reasons, a different approach has to be chosen. The idea is to use a detailed and highly accurate digital map (Road Work Map), which contains the positions of all important elements of a road construction site, for example, road markings, crash barriers, and pylons/beacons. This additional information is then used to fully reconstruct the environmental model of the vehicle, if parts of the surrounding cannot be measured with on-board sensors. For a robust detection, active sensors, such as lasers scanners or radar, are suitable as they are less susceptible to weather conditions. The laser scanner is not only able to detect other vehicles but also beacons, traffic pylons, and guard rails, which are commonly used at work zones. This data then is fused with the data of a video camera, a position estimate derived from GPS, and the detailed Road Work Map described above. Therefore, robust and reliable data can be supplied for assistance systems such as lane keeping.

Controlling a heavy-duty vehicle laterally is a demanding task [13–15]. Many systems require a look-ahead capability to enable a robust control. This can be easily provided, as the future road course can be extracted from the proposed Road Work Map.

The concept described above is divided into two parts (Fig. 1). First, the highly accurate map has to be created. For this, a measurement vehicle is equipped



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 1
Overview over the modules for creating a highly accurate Road Work Map (*left*). Proposed system setup, which uses the Road Work Map to supply reliable data for advanced driver assistance systems (*right*)

with additional sensors. Each time when a new road construction site is set up, the vehicle will drive through the scenery and automatically detect and map all important elements. Here, post-processing and the use of additional and more accurate hardware are possible. This generated Road Work Map is transferred to facilities at the entrance of the road work zone.

The second part deals with the application for all vehicles driving through the construction site. It is assumed that cars and trucks receive a current map from the work zone via wireless communication. If the vehicles are equipped with a future sensor, for example, a laser scanner, they can match their current environmental model to the map and thus profit from additional information for assistance systems such as lane keeping assist.

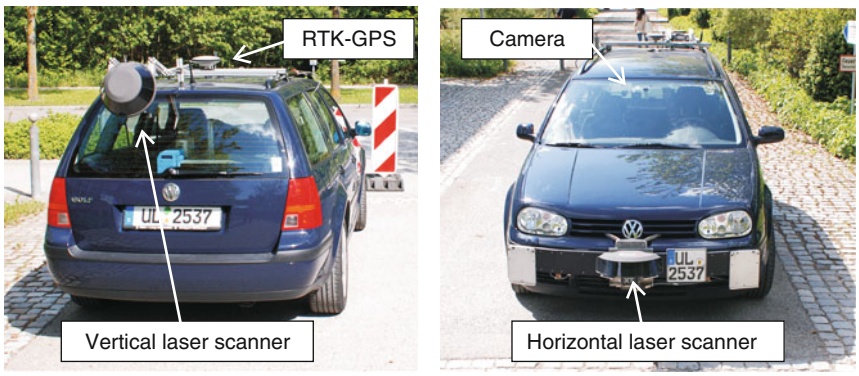
The automatic generation of the Road Work Map will be presented in the following sections.

Sensors and Measurement Vehicle

The measurement vehicle is a standard car of series production, but additionally equipped with several sensors for environmental perception (Fig. 2). A front-facing gray-scale camera with a resolution of 640 × 480 pixels is used for the detection of road markings. A wide-angle lens allows the measurement of markings not only from the lanes of the host vehicle, but also from neighboring lanes.

Two Ibeo ALASCA XT laser scanners [16] are used for acquiring a horizontal and a vertical distance profile of the surrounding area. This type of sensor sends out a rotating laser beam and detects the echo reflected on obstacles. With the speed of light and the time of flight, the distance to an object can be derived. The laser scanner has an angular resolution between 0.125° and 0.5°, depending on the angle and scanning frequency. The total horizontal field of view is up to 240°; at this vehicle this is limited to 180° due to constraints of the housing and mounting position. The laser scanner provides four layers with a vertical angle shifted by 0.8°, resulting in a vertical opening angle of 3.2°. That feature enables the laser scanner to detect objects even in cases where the vehicle pitches, thus making tracking more robust. The detection range is up to 200 m, depending on the reflectivity. The good distance resolution of 4 cm enables a highly accurate detection and mapping. The front-facing laser scanner is used to detect beacons and traffic pylons as well as detecting the free space in front of the vehicle. At the rear of the vehicle, another laser scanner is mounted to acquire a vertical distance profile. This is used to detect and map safety barriers such as guard rails and Jersey barriers, which limit the road to the left and right.

Besides the novel environmental sensors, series sensors are used for estimating the position and movement of the host vehicle. Wheel encoders and a MEMS (micro-electro-mechanical systems) yaw rate sensor of a modern ESP (Electronic Stability Program) system are used for calculating the relative host vehicle movement between successive measurement frames. A standard DGPS (Differential Global Positioning System) provides global position information of about 3–10 m accuracy, depending on the measurement conditions of



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 2
The measurement vehicle with two laser scanners, a video camera, and a highly accurate RTK-GPS

the satellites. This type of sensor is already part of series cars equipped with an on-board navigation system.

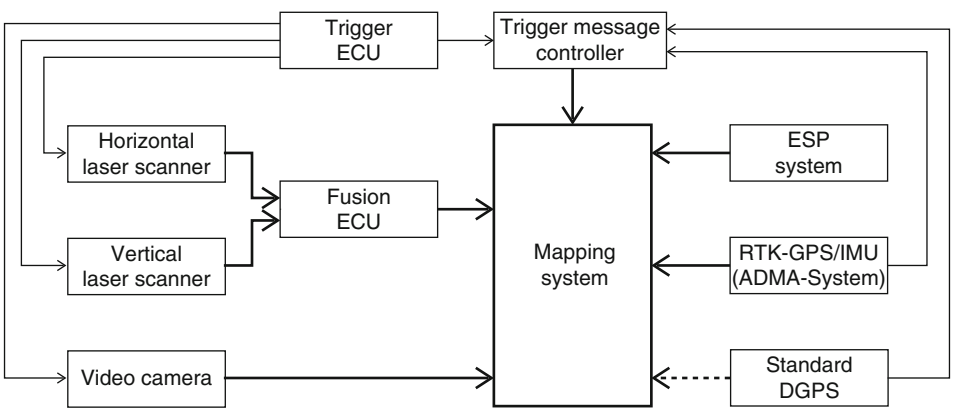
For generating the Road Work Map, a highly accurate position of the measurement vehicle must be known in a global coordinate system, for example, WGS84 [17] or UTM [18]. This position is acquired using a RTK-GPS (Real Time Kinematic GPS), which additionally uses online correction data of the land survey administration, thus enabling a horizontal position accuracy of 1 cm (RMS). This system not only uses GPS, but also GLONASS, the Russian Global Navigation Satellite System (GNSS). High accuracies can only be achieved under good measurement conditions. In cases where no or only a little number of satellites is visible (e.g., in tunnels or in street canyons), an IMU (inertial measurement unit), which contains highly accurate yaw velocity and acceleration sensors, is used. Both data is fused to obtain best position estimation (ADMA: Automotive Dynamic Motion Analyzer, [19]).

All data of position sensors and environmental sensors must be transformed to a global coordinate system. For that reason, the mounting position and alignment of each sensor relative to the vehicle must be known. Laser scanner and video camera have to be calibrated [20] and the position of the ADMA has to be known.

Besides this, timing of the sensors is similarly important on a moving platform like the measurement vehicle. Typical speeds at road work zones on highways are 80 km/h. This means, that a timing inconsistency of 5 ms already accounts for an error of over 11 cm. For that reason, laser scanner and video camera are hardware-triggered; the measurement frequency is set to 12.5 Hz. The highly accurate GPS system provides a trigger pulse every 20 ms (50 Hz). Both pulses are input to a trigger message controller which generates a time stamp message on the CAN bus (see Fig. 3). The data of the ESP is already present at the CAN bus. All of these messages are used to align sensor data from different sources to a common time reference.

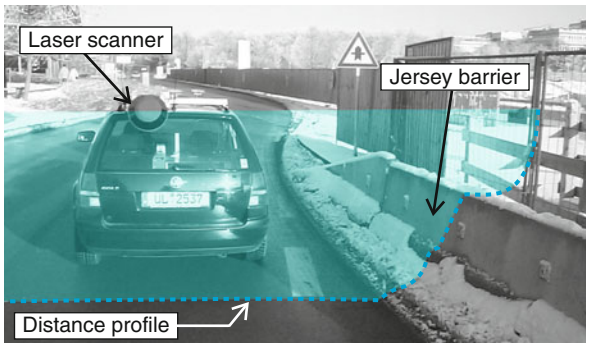
Mapping of Safety Barriers

High speeds at highways and freeways even at road construction sites demand a mechanically robust protection system both for workers as well as motorists. Often, the traffic routing is changed significantly. A lane may be omitted, the width will be reduced and even one lane may be routed to the side of the opposing traffic. This poses a potential danger for all road users. For that reason, different safety barriers have been developed [21]. Crash barriers (guard rails) can be found on almost every dual carriageway. Especially for road construction sites, removable barriers are used. A commonly used type is the Jersey barrier, which is made of concrete and exists in different sizes. But steel barriers are also frequently used. Safety barriers in general serve different needs, which may contradict each other. On the one hand, these objects should have a preferably small profile to offer enough



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 3
Schematic of sensors and modules of the measurement vehicle. All important sensors have trigger capability to provide accurate timing information

space for the actual traffic and the area of work. On the other hand, these barriers have to withstand impacts of heavy duty vehicles, so that these trucks can be kept within their lane in the event of a crash. Additionally, the barriers are designed to absorb some amount of crash energy, so that vehicles do not directly bounce off the wall. There are regulations which specify the physical characteristics in the event of a crash, subject to the total speed of the vehicle, its tonnage, and the angle under which it hits the barrier. Depending on the maximum speed limit and other operation purposes, different rules may apply [22]. The actual shape or material, however, is not regulated. For that reason, a detection system has to deal with numerous different outlines [21].

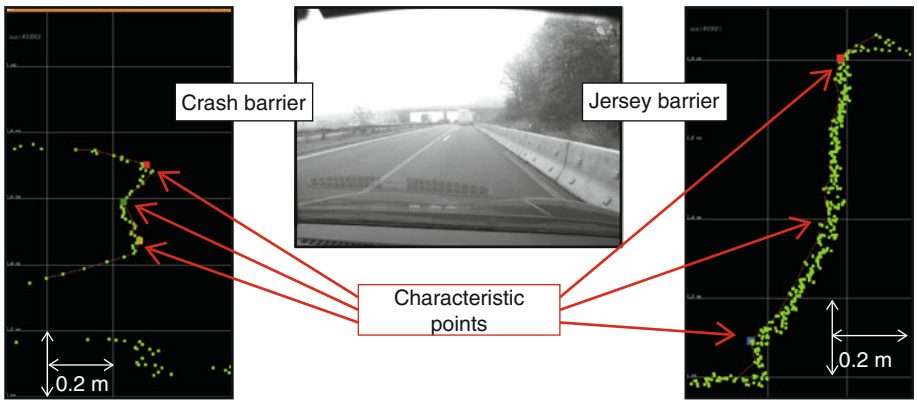


Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 4
Measurement principle of a vertically mounted laser scanner. A vertical distance profile enables the detection of barriers, which are located *left* and *right* of the vehicle

A vertically scanning laser scanner mounted at the rear of the vehicle has been chosen as sensor setup for detecting and mapping safety barriers. The measurement principle is shown in Fig. 4. Because of the high angular resolution and the four layers, a detailed outline can be obtained. This data undergoes further analysis to distinguish between background (e.g., vegetation) and actual work zone infrastructure.

As a first step, the data is filtered to reduce measurement noise and to sharpen the contour [23]. A third-order Gaussian noise filter is applied on all raw laser scan points. On average roads, the mapping vehicle is subject to slight rolling, which distorts the measurement. Since the laser scanner acquires a distance profile of the road directly behind the rear of the vehicle, a regression line can be fit into the data. It can be assumed that the area which the vehicle is driving on is even. Therefore, a compensation of the roll angle is possible.

The structures of interest extend in vertical direction, as is illustrated in Fig. 5. That fact is exploited to extract suitable candidates for safety barriers. A histogram is applied on the data, where several vertical slots accumulate the number of scan points in this region. The distance between successive scan points



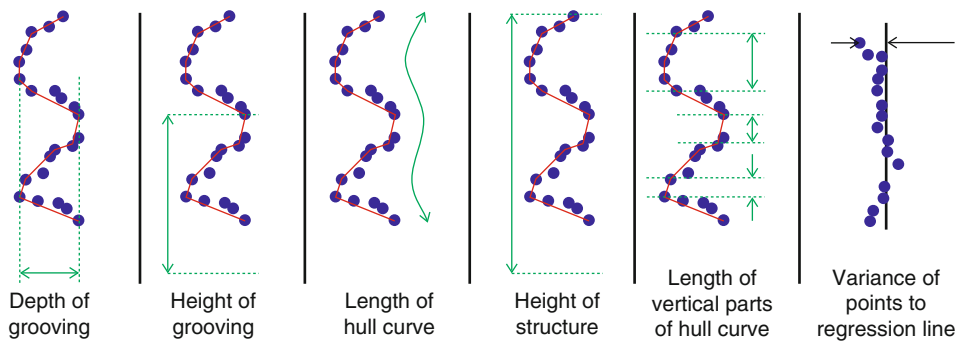
Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 5
Real scan data example of a vertical scan. A guard rail is located on the *left* side, whereas a Jersey barrier is on the *right* side. As a result of the detection algorithm, three characteristic points can be extracted to incorporate them into the Road Work Map

increases in further distances from the vehicle due to the radial measurement principle of the scanner. This effect can be compensated before the accumulation is performed by applying different weights on the points based on the geometry [24]. The bins with a higher rating are regarded as valid candidates and are further examined.

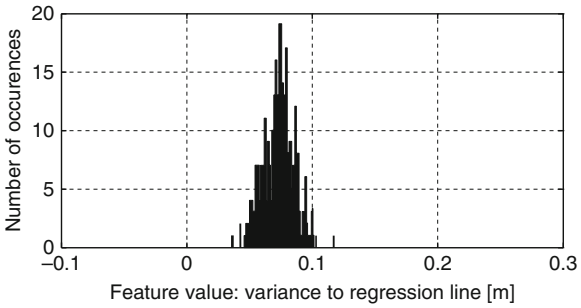
Automatically detecting safety barriers requires distinguishing between background and actual infrastructure. Grass and bushes often grow underneath and around guard rails, which is also measured by the laser scanner. The Road Work Map, however, should not include background data, as this is not the border of the road. Preliminary tests have shown, that a rule-based separation of infrastructure and background only yields low detection rates. For that reason, a more sophisticated approach had to be chosen. As mentioned above, there is a variety of different safety barrier types. Extensive tests in Germany have shown that all different barriers can be grouped into a few different classes, which can be seen in Fig. 9. A similar analysis can be performed for other regions and barrier types. These types may not be totally the same structure, but they have a similar sensor reading, hence an analog outline. Separating between different classes leads to the use of a Bayesian classifier.

For making use of a Bayesian classifier, features have to be found, which best describe the different classes. The vertical distance profile of the laser scanner is processed to find a representation of the typical outline of safety barriers. As the scan points are subject to noise, a hull curve is fit into the measurement points. This hull curve is designed to represent the outer hull (closest to the vehicle), as a conservative approximation of the barrier with respect to the application. The hull curve is the basis for further feature extraction algorithms, which are based on the geometry of the structure. One obvious useful feature is the height of the barrier. Tall Jersey barriers reach approximately 1.2 m whereas guard rails have a typical height over ground of 0.8 m. Two other features are directly derived from the above-mentioned hull curve. One feature is the total length of the contour, the other one is the length of all vertical parts of the hull curve. Based on an angle interval criterion only those segments are accumulated, which are regarded as approximately vertical. Another two features are derived from the shape of guard rails. These structures show a typical grooving in the middle of the metal crash barrier. The height of this grooving is taken as one feature; the calculated width (horizontal distance of closest and furthest point to the vehicle) is taken as the other feature. For separating artificial structures like Jersey barriers from background vegetation a sixth feature is extracted: a regression line is fit into the vertical scan points. The variance of the horizontal distance of the points is regarded as a means for the smoothness of the structure. An illustration of all features can be seen in Fig. 6.

The features described above are evaluated for all different kinds of safety barriers. Several hundreds of data frames have been labeled manually to gain the typical distribution of the feature values. The result for one class and one feature value is exemplarily shown in Fig. 7. It can be seen that the class-conditional



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 6
Several feature values can be calculated from the vertical scan based on the geometry of the structure. These features are used as an input for the Bayesian classifier



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 7
Number of occurrences of the feature values ("variance of points to regression line") for one barrier type (metal crash barrier)

probability density function $p(x|C_i)$ for a feature x and class C_i corresponds to a Gaussian distribution.

Hence, the density functions can be approximated by the mean μ and variance σ^2 of the features for every class. This fact motivates the use of a Bayesian classifier [25]. It is assumed, that all types of barriers (classes C_i) have equal prior probabilities $P(C_i)$. $P(C_i|x)$ is the probability that class i is true for the given feature value x (posterior probability) and should be calculated based on the measurement of all feature values and the predetermined class-conditional probability density functions.

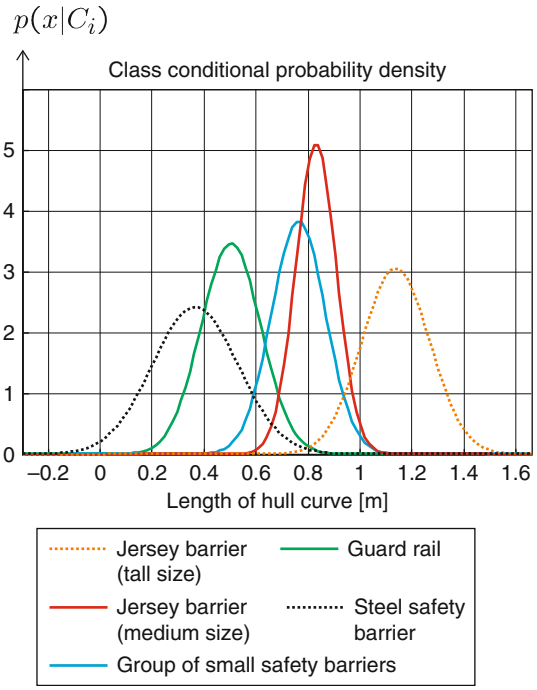
This can be calculated for a one-dimensional feature and N classes:

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{\sum_{j=1}^{N} p(x|C_j)P(C_j)}$$

The class-conditional probability density functions of all types of barriers for the feature "length of hull curve" can be seen in Fig. 8. For the six-dimensional ($d = 6$) feature vector $\underline{x}$ (mean vector $\underline{\mu}$ and covariance matrix $\underline{K}$), the class-conditional probability density is:

$$P(\underline{x}|C_i) = \frac{\exp\left(-0.5\left(\underline{x}-\underline{\mu}\right)^T \underline{K}^{-1}\left(\underline{x}-\underline{\mu}\right)\right)}{(2\pi)^{0.5d}\sqrt{|\underline{K}|}}$$

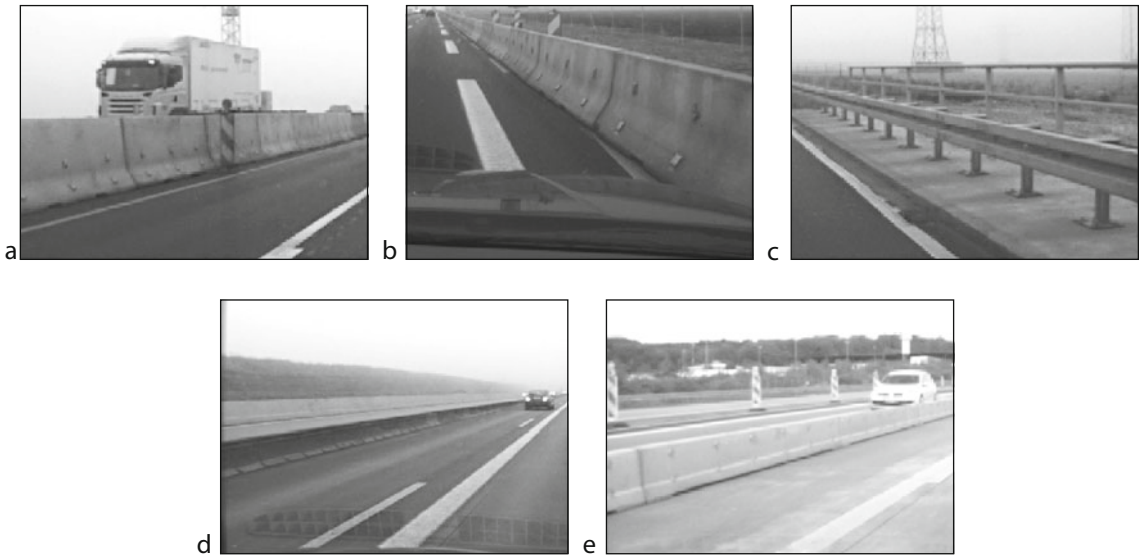
The extension to the multidimensional case results in a most probable class (type of barrier) for the given feature vector. This does not yet include the case, where the candidate structure is measured in the background.



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 8
Class conditional probability density functions for five different types of barriers for the feature "length of hull curve"

Therefore, only those measurements are regarded as valid, which lie in between the 2σ interval of its mean. This value has been found as best compromise between detection rate and the rate of non-detected safety barriers. Approximately 3,300 measurements have been labeled manually for evaluating the performance of the classifier. The overall detection rate is 97%, which indicates the differentiation between background and safety barriers. This number includes some misclassifications between different classes, that is, a Jersey barrier might have been mistaken as steel barrier (Fig. 9). Regarding correct class assignment, the correct classification rate is about 95%. Non-detected elements occur in 3% of the measurements, where the false positive rate (background erroneously detected as safety barrier) is only 2%.

After the classification of all candidates, a suitable representation of the structure is calculated. The top most and bottom point is extracted, as well as the position of the typical grooving for barriers which are



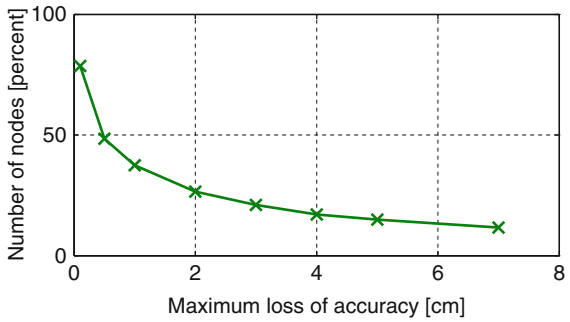
Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 9
Examples of all five different types of barriers used for the classification algorithm: (**a**) tall Jersey barrier, (**b**) medium-size Jersey barrier, (**c**) standard guard rail, (**d**) steel safety barrier, (**e**) example of a group of different small safety barriers

classified as guard rails. For all other barriers, top and bottom points are derived as well, but additionally, the position of the bend of Jersey barriers is extracted. All safety barriers are now represented by three characteristic points (see Fig. 5). This enables a lower amount of data to be stored than using the whole hull curve. But on the other hand, the structure is still represented with high accuracy.

The above-mentioned process has been performed on a single frame basis. As the detection of safety barriers is meant for mapping, the connection of successive measurement points can be done with post-processing calculations. First of all, all measurement points are transformed from the local sensor coordinate system to a global frame using the position of the vehicle in global coordinates derived from the ADMA system. Then, all nodes which originate from the same class are grouped to segments. An outlier detection based on distance criteria removes false points of the classification process. All data of one segment is interpolated by a smoothing spline [26, 27]. On the one hand, this helps to bridge gaps between non measured areas (due to e.g., occlusion) and on the other hand, additional smoothing of noisy data is possible. A smoothing spline can be regarded as a combination of an interpolating spline, which connects all nodes by a piecewise continuous polynomial, and a regression line. A regression line minimizes the squared distance of all nodes to the function, where the spline, in contrast, minimizes the curvature of the function. The trade-off between those characteristics can be chosen by the smoothing factor λ. $S_i(x)$ is the polynomial in segment i and y_i the node with an assumed standard deviation of σ_i. Minimizing the function L gives the desired smoothing spline:

$$L = \lambda \sum_{i=0}^{n} \left(\frac{y_i - S_i(x)}{\sigma_i} \right)^2 + (1-\lambda) \int_{x_0}^{x_n} \{S_i''(x)\}^2 dx.$$

As the Road Work Map is designed to be transferred via wireless communication, the amount of data should be as low as possible. As the spline offers a very flexible representation of arbitrary curves, the user is able to omit nodes without losing too much accuracy. A maximum distance between the original node and the spline reconstruction of the barrier with less supporting points can be set to a value which is determined by the application. The exemplary relationship between the maximum loss of accuracy and the number of nodes of a sample barrier is shown in Fig. 10.



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 10
Exemplary relationship between the parameter "maximum loss of accuracy" and the number of nodes necessary for representing the barrier

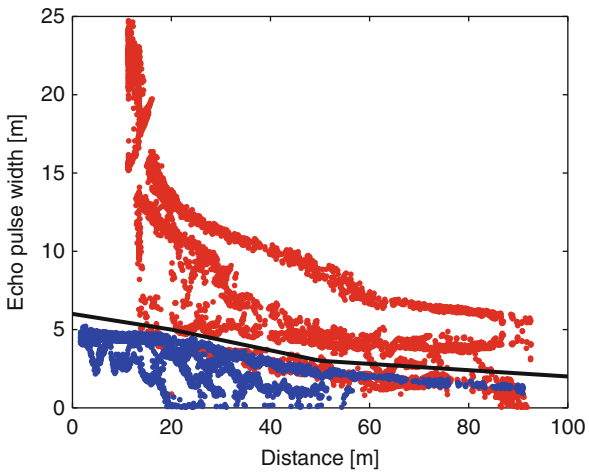
Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 11
Two infrastructure elements frequently used at road construction sites: beacon or guiding reflector post (*left*), traffic pylon or cone (*right*)

Besides the detection rate, the overall accuracy which can be achieved with that concept is of interest. For evaluating the preciseness of the algorithms mentioned above, the exact position of a guard rail has been manually derived by a static RTK-GPS measurement, which yields to a position accuracy of about 3 mm (1 σ interval) [28]. This is compared to the results of the automatic detection and mapping process. The evaluation shows a mean squared error of 7.1 cm and a standard deviation of 5.7 cm. This means that the proposed set-up produces results accurate enough for the generation of a detailed Road Work Map.

Detection of Beacons and Traffic Pylons

At the beginning and ending of road construction sites traffic pylons and guiding reflector posts (beacons) are often used (Fig. 11). They commonly mark the lead-in and exit tapers and guide drivers when traffic routing is changed. Pylons and Beacons are used especially for works of short duration [22], for example, when new road markings are applied on the road, as this is less time consuming than erecting safety barriers.

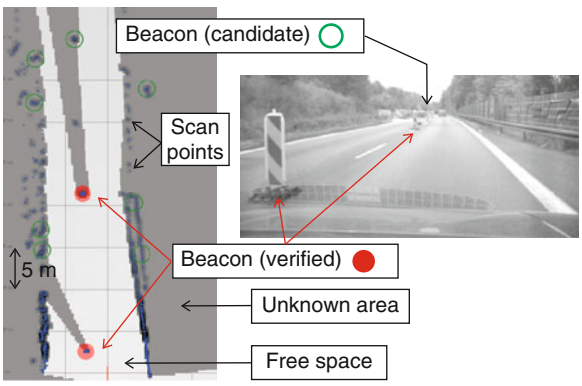
Both elements must have a reflecting surface. This fact is exploited when pylons and beacons should be detected with a laser scanner. The laser sensor not only measures the distance to objects but also the so called echo pulse width, which is proportional to the amount of energy returned from targets. The total amount of energy decreases with the distance of objects from the sensor. By comparing the data of reflecting and non-reflecting materials for different distances, a characteristic can be deduced, which is shown in Fig. 12. This characteristic enables the classification of each scan point to be reflector or non-reflector and is only slightly influenced by weather conditions.



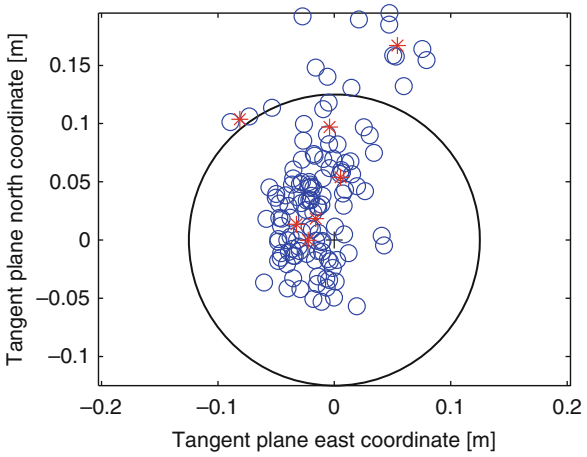
Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 12
Distinguishing between reflecting and nonreflecting scan points based on the echo pulse width and the distance from the sensor. Scan points which originate from reflecting surfaces are drawn in *red*; points from nonreflecting surfaces are colored in *blue*. The *black* line shows the limiting curve which is used in the algorithm

Besides the reflector attribute, the size of the measured objects is taken into account as well. For extracting suitable beacon candidates, a measurement grid of the laser scan is used. This concept has been introduced in the field of robotics [29] and is now also used for automotive applications [30]. The area in front of the vehicle is segmented into a grid of squares with, for example, 20 cm edge length and initialized with probability value 0.5. For each scan point within one cell, the value is increased, depending on its distance and radial position to account for measurement noise. The area between the sensor and a measurement is assumed to be unoccupied, so the value of the cells on the way of the laser beam is decreased. Due to noise, this is also done distance dependent. The grid cell values can be visualized in grey scale color, where black means occupancy of 1 and white means 0 probability of occupancy (see also Fig. 13). As a result, the grid contains dark areas which are presumably occupied and hence denote the position of obstacles. Other grid cells have lighter color which means they are assumed to be on free space or the region is unknown, that is, the value remains 0.5. Based on this measurement grid, a contour extraction algorithm is applied to find the position of free standing objects with a certain size. If these objects additionally contain scan points marked as reflectors, they are regarded as candidates for pylons and beacons. After the object generation based on single frame, a tracking over time is applied. The position of all candidates of one measurement time is transformed into the next frame by compensating the host vehicle movement, as the pylons are assumed to be stationary. An association algorithm finds correspondences between last and current frame. In this way, random detections in the background can be filtered out, if no new measurements are found in this area. A track confirmation finally verifies pylon and beacon candidates. These positions can either be provided for online use while driving, or for incorporating the objects into the Road Work Map.

For assessing the preciseness of the detection and tracking algorithm, the position of some beacons and pylons has been determined manually with a static RTK-GPS measurement. The accuracy of this reference positions is approximately 3 mm (1 σ interval). This is compared to the automatic mapping process. The mean squared error evaluates to 7.3 – 7.7 cm for beacons and pylons respectively (see Fig. 14). With respect to the sensor resolution, this is a good value.



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 13
Detection of beacons using a front-facing laser scanner. Exploiting both size and reflectivity of objects enables the extraction of suitable candidates. A tracking algorithm verifies true beacons



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 14
Accuracy of the mapping algorithm for traffic pylons. The true position (reference) is shown by a *black* circle with a diameter of 25 cm, which is the size of the bottom of the pylon. Each single measurement is drawn as a *blue* circle. The tracked position, which would be incorporated into the map, is shown as *red* star

Detection and Mapping of Road Markings

Road markings are used for guiding drivers on all major roads. They instruct motorists, if a lane change is possible or not. At lead-in and exit tapers, they are especially important for a safe and smooth guidance of all vehicles. At work zones the original marking (e.g., white color) may be invalidated by markings of another color (e.g., yellow), while the primary lines still remain on the road. That is the reason why lane recognition in road construction sites is such a demanding task. The Road Work Map described above should provide an interpreted description of the whole environment, including road markings.

Lane recognition with video cameras already has a long history in research. For well-structured scenarios such as highways and freeways, assistance systems for lane departure warning already exist on the market. The specific needs of such systems for road work zones, however, are not yet addressed. Current lane detection approaches offer a wide variety of concepts [31]. The first step usually is a feature extraction, followed by road modeling and finally a temporal tracking. These steps can be performed either based on monocular [32] or stereo [33] images. One of the most common concepts for feature extraction is to find edges in the image caused by a light-to-dark or dark-to-light transition of white or yellow markings on a darker road surface. Other approaches use color information or try to find areas of similar texture [34]. A template matching such as the Hough transform [32] can be applied to fit the data to a predefined model. Depending on the type of road which is mostly addressed by the application, an adapted mathematical curve is fit to the data. Commonly used are parabolic [31] or cubic curves, the assumption of constant curvature [32], or approximations for clothoids [35, 36]. The assumption of a flat world [31] may hold for many applications; otherwise, the road has to be modeled three dimensional [37]. A temporal filter such as a Kalman filter or particle filter can be applied to gain additional robustness against single frame detections [32, 34, 38].

The road markings in work zones have to be treated in a special case. The following paragraphs will describe all algorithms and adaptations for creating a highly accurate environmental model of road markings for the purpose of mapping. All lane markings should be detected at one single drive-through; therefore a video camera with wide-angle lens is used. This results in a detection range of about 20 – 35 m which is enough for mapping. Additionally, modeling a flat world in this range is a valid assumption on highways and freeways.

For edge detection, a Canny filter [39] has been chosen. This commonly used algorithm consists of multiple stages, among them are a Gaussian filter for noise reduction, an operator for determining magnitude and direction of the edge, and a non-maximum suppression. Road markings are assumed to have a left (dark to light transition) and right (light to dark transition) edge. This is exploited for finding appropriate pairs of edge points. The magnitude G and direction θ of the edge is calculated with the horizontal (x) and vertical (y) Sobel operator [39] on the image $\underline{I}$:

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * \underline{I}$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \underline{I}$$

$$G = \sqrt{G_x^2 + G_y^2}$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

Due to image noise and blur there may be several edges in close distance. A local maximum gradient search finds the most probable position of the edge. These points are projected from image domain to vehicle coordinate system using a flat world assumption. The distance between left and right edge is the measured width of the road marking. Edge point pairs are only considered valid, if the distance matches the regulations for road marking widths [40] including some noise margin. Finally, only a little number of false detections remains.

After that, those point pairs which originate from the same road marking have to be associated. A linear extrapolation in the driving direction is calculated based on a least-square fit to associate adjacent measurement nodes. Point pairs are associated, if they have

a similar width and an appropriate distance to the preceding point pair. The measurement of lengths and widths is subject to noise. These errors can have several reasons and must be considered. The quantization through pixels results in different error values in the road plane, that is, they will be little in near distance from the vehicle and bigger in long distances. Additionally, pitching and rolling of the vehicle will result in errors because of the projection from image to flat road surface. These influences are taken into account before association. Point pairs, which cannot be associated are regarded as outliers and are neglected.

For separating one lane from the neighboring lane, dashed road markings are commonly used. The line length usually is 6 m, the gap between two line segments can be either 6 or 12 m [40, 41]. The detection algorithm only finds edges on the actual markings, so it is desirable to associate adjacent lines. Therefore, a higher order regression polynomial f_M is fit into the whole data of each line segment:

$$f_M(t, \underline{a}) = a_0 + a_1 t + \ldots + a_n t^n$$

where $a_0, \ldots, a_n$ are the polynomial coefficients and n the order of the polynomial, which can be adapted depending on the number of nodes. Finding the optimal parameters $\hat{\underline{a}}$ on a set of m data points y_i means minimizing the squared error sum S[42]:

$$\min_a S = \min_a \sum_{i=1}^{m} (y_i - f_M(t_i, \underline{a}))^2$$

This problem can be rewritten:

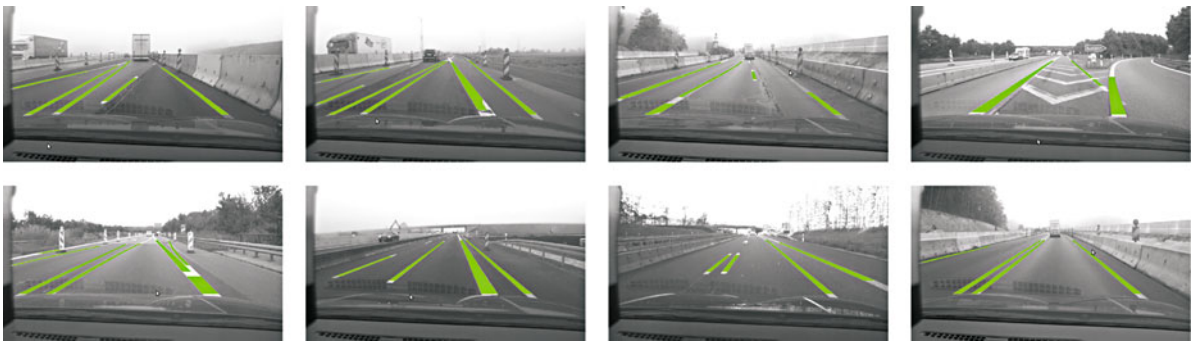
$$\begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_m & t_m^2 & \cdots & t_m^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$$\underline{B\hat{a}} = \underline{y}$$

$$\hat{\underline{a}} = (\underline{B}^T \underline{B})^{-1} \underline{B}^T \underline{y}$$

The results are two polynomials which can be extrapolated and compared via distance criteria. If they overlap, these two lines are assumed to be part of the same lane marking. This method also helps to bridge gaps between non-detected segments (e.g., due to occlusion) of road markings. As road markings can have a curvature, especially when the traffic routing changes, this method assures a high flexibility for arbitrary curvatures within the field of view of the camera. Modeling curves for longer distances is done by splines in the post-processing step.

The detections based on single frames (illustrated in Fig. 15) are handed over to several post-processing stages. First, all data of the same road marking, which has been measured several times in different time steps has to be associated. This is based on distance criteria under consideration of the regulations for lane markings. As a result, there will be a lot of nodes of the same line, depending on the performance of the detection algorithm. For further processing, these nodes need to be clustered and, if necessary, the number of nodes can



Driver Assistance Systems, Automatic Detection and Site Mapping. Figure 15
Examples of the road marking detection algorithm. The images show the performance for different lighting and surface conditions

be reduced. That is done by an algorithm inspired by the k-means clustering [43] with fixed cluster size, which preserves enough data points in areas with sparse number of points. After that, an outlier detection based on geometric parameters is applied. Background and beacons, which sometimes produce erroneous detections, can thus be removed, as they appear only in a limited number of frames and in inconsistent directions. Similar to the post-processing steps of the safety barrier mapping, a suppression of noise is achieved through the calculation of a smoothing spline and the number of nodes is reduced by an iterative process with predefined maximum error.

Highly Accurate Road Work Map

All three different elements of road construction sites can now be incorporated into the map. Beacons and traffic pylons have been measured by a front-facing laser scanner. Safety barriers are detected by a vertically scanning laser sensor and road markings are identified with a video camera. This data can now be interpreted with respect to valid and invalid lane markings. The idea is that vehicles which approach the work zones will receive a current and detailed Road Work Map. Usually, cars will not be equipped with highly accurate RTK-GPS sensors, but with standard DGPS already used for navigation applications. A matching algorithm aligns the map data with the current environmental perception, thus enabling a more accurate and extended model of the construction site. This is used as input for driver assistance systems such as lane keeping support.

Future Directions

Increasing traffic density makes driving more and more challenging. Assistance systems help to support motorists in their basic tasks. In the future, semiautomated vehicles are not completely unrealistic. Driving long distances on well-structured environments like freeways could partly be handed over from human to intelligent vehicles equipped with longitudinal and lateral control. More complex scenarios like urban traffic with pedestrians and cyclists sharing the same traffic area will still require human supervision, but there are already concepts for supporting the driver in more and more tasks.

The proposed concept of transferring maps to approaching cars and trucks requires infrastructure-to-vehicle communication. All major vehicle manufactures are involved in the research of applications for vehicle-to-vehicle or infrastructure-to-vehicle communication applications. The technical details are not yet fully specified, but the development shows a promising progress. This would enable the use of highly accurate infrastructure maps for many different future driver assistance systems [44]. These maps could also be extended by dynamic objects, such as pedestrians at intersections in urban areas.

Bibliography

1. Bureau of Transportation Statistics (2010) U.S. Department of Transportation, National Transportation Statistics, Washington
2. European Commission (ed) (2009) Panorama of transport, office for official publications of the European communities, Luxembourg
3. Helling A (1997) Transportation and economic development: a review. In: Little RG(ed) Public works management and policy, vol 2 (1). Sage, Thousand Oaks, CA, USA, pp 79–93
4. Benekohal R, Hashmi A, Orloski R (1993) Drivers' opinion on work zone traffic control. In: Transportation quarterly, 47 (1). ENO Foundation, Eno Transportation Foundation, Washington, DC, USA, pp 19–38
5. DEKRA (ed) (2009) Road works cause anxiety – Baustellen sorgen fuer Nervenflattern, DEKRA press and information, Stuttgart
6. Khattak Asad J, Khattak Aemal J, Council FM (2002) Effects of work zone presence on injury and non-injury crashes. In: Accident analysis and prevention, vol 34(1). England, Elsevier Science Ltd, Amsterdam, Netherlands, pp 19–29
7. ARROWS Consortium (1999) Advanced research on road work zone safety standards in Europe (arrows), road work zone safety, practical handbook, background report. National Technical University Athens
8. Board of Trustees for Traffic Safety (ed) (2007) Hazard area freeway road works? – Gefahrenstelle Autobahnbaustelle? Kuratorium für Verkehrssicherheit – Board of trustees for traffic safety, Austria
9. Twisk D, Mesken J (2007) PREVENT – Education to improve road safety around work zones. In: Larsson TJ Safety science monitor, vol 11(2). Perth, Australia
10. Wang J, Hughes WE, Council FM, Paniati JF (1996) Investigation of highway work zone crashes: what we know and what we don't know. In: Transportation research record 1529. TRB, National Research Council, Washington
11. Marsden G, McDonald M, Brackstone M (2000) Towards an understanding of adaptive cruise control. In: Transportation research part C. Elsevier Science, Pergamon

12. Bishop R (2000) A survey of intelligent vehicle applications worldwide. In: Proceedings of the IEEE intelligent vehicles symposium, Dearborn, 2000
13. Martini S, Murdocco V (2003) Lateral control of tractor-trailer vehicles. In: Proceedings of 2003 IEEE Conference on Control Applications, CCA
14. Wang J-Y (1999) Robust H∞ Lateral control of heavy-duty vehicles in automated highway system. In: Proceedings of the American control conference, San Diego
15. Fritz H (1999) Longitudinal and lateral control of heavy duty trucks for automated vehicle following in mixed traffic: experimental result from the CHAUFFEUR project. In: Proceedings of the 1999 IEEE international conference on control applications. Kohala Coast-Island of Hawai'i, Hawai'i
16. Ibeo Automobile Sensor GmbH (2006) ALASCA user manual, Hamburg
17. European Organisation for the Safety of Air Navigation (1998) WGS 84 implementation manual, European organisation for the safety of air navigation, Brussels, Belgium and Institute of Geodesy and Navigation University FAF Munich, Germany
18. Osborne P (2008) The mercator projections, edinburgh, http://mercator.myzen.co.uk/
19. GeneSys GmbH (2010) ADMA user manual, Offenburg
20. Kaempchen N (2007) Feature-level fusion of laser scanner and video data for advanced driver assistance systems. Ph.D. dissertation, University of Ulm, Ulm, Germany
21. Knoll E (ed) (2004) Der elsner – handbook for traffic and transportation. Otto Elsner Verlagsgesellschaft mbH & Co. KG, Dieburg, Germany
22. German Ministry of Transport (Bundesministerium für Verkehr, Bau- und Wohnungswesen, Abteilung Straßenbau (BMVBS)) (1995) Safety instructions for road construction sites – Richtlinien für die Sicherung von Arbeitsstellen an Straßen, Bonn, Germany
23. Wimmer A, Weiss T, Flögel F, Dietmayer K (2009) Automatic detection and classification of safety barriers in road construction sites using a laser scanner, In: IEEE proceedings of intelligent vehicles symposium, Xi'an, China
24. Weiss T, Dietmayer K (2007) Automatic detection of traffic infrastructure objects for the rapid generation of detailed digital maps using laser scanners. In: Proceedings of the 2007 IEEE intelligent vehicles symposium, Istanbul, Turkey
25. Duda RO, Hart PE, Stork DG (2000) Pattern classification. Wiley, New York
26. Reinsch CH (1967) Smoothing by spline functions. In: Reinsch CH (ed) Numerische mathematik, vol 10. Oxford University Press, New York, pp 177–183
27. Pollock D S G (1993) Smoothing with cubic splines, Department of Economics, Queen Mary College, University of London
28. NovAtel Inc (2007) Datasheet for receiver OEMV-3. NovAtel Inc, Canada
29. Thrun S, Fox D, Burgard W (2005) Probabilistic robotics. MIT Press, Cambridge
30. Weiss T, Schiele B, Dietmayer K (2007) Robust driving path detection in Urban and highway scenarios using a laser scanner and online occupancy grids. In: IEEE proceedings of intelligent vehicles symposium, Istanbul, Turkey
31. McCall JC, Trivedi M (2006) Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. In: IEEE transactions on intelligent transportation systems, Piscataway, NJ, USA
32. Taylor C, Košecká J, Blasi R, Malik J (1999) A comparative study of vision-based lateral control strategies for autonomous highway driving. I J Robotics Res, Sage Publications, Thousand Oaks, CA, USA
33. Bertozzi M, Broggi A (1998) GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection. In: IEEE transactions on image processing, Piscataway, NJ, USA
34. Apostoloff N, Zelinsky A (2003) Robust vision based lane tracking using multiple cues and particle filtering. In: Proceedings of IEEE intelligent vehicles symposium, Piscataway, NJ, USA
35. Dickmanns ED, Mysliwetz BD (1992) Recursive 3-D road and relative ego-state recognition. In: IEEE transactions on pattern analysis and machine intelligence, vol 14(2), IEEE Computer Society, Washington, DC, USA, pp 199–213
36. Gumpp T, Nienhüser D, Liebig R, Zöllner JM (2009) Recognition and tracking of temporary lanes in motorway construction sites. In: IEEE proceedings of intelligent vehicles symposium, Xi'an, China, IEEE, Piscataway, NJ, USA
37. Nedevschi S, Schmidt R, Graf T, Danescu R (2004) 3D Lane detection system based on stereovision. In: IEEE intelligent transportation systems conference, Piscataway, NJ, USA
38. Loose H, Franke U, Stiller C (2009) Kalman particle filter for lane recognition on rural roads. In: IEEE proceedings of intelligent vehicles symposium, Xi'an, China, Piscataway, NJ, USA
39. Jähne B (2005) Digitale bildverarbeitung. Springer, Berlin
40. Bundesministerium für Verkehr (ed) (1993), Instructions for the construction of road markings – Richtlinien für die Markierung von Straßen, Bundesministerium für Verkehr – German Ministry of Transport, Köln
41. Research Association for Road and Transportation (ed) (1996) Instructions for the construction of roads – richtlinien für die anlage von Straßen, Forschungsgesellschaft für Straßen- und Verkehrswesen – Research association for road and transportation, Köln
42. Ljung L (1999) System identification: theory for the user. Prentice Hall, New Jersey
43. MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of fifth Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press
44. Blervaque V, Mezger K, Beuk L, Loewenau J (2006) ADAS horizon – how digital maps can contribute to road safety. In: Proceedings of advanced microsystems for automotive applications Berlin

Driver Behavior at Intersections

Toshihisa Sato, Motoyuki Akamatsu
Human Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba-shi, Ibaraki, Japan

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Behavioral Data Collection in an Actual Road Environment
Analysis of Data Distribution Based on Traffic and Road Environments
Modeling of Driver's Preparatory Behavior Before Making Driver's Side Turn
Future Directions
Bibliography

Glossary

Driving simulator An indoor system providing a multisensory environment for a driver to perceive and control virtual vehicle movements. A standard simulator has a vehicle cockpit; a visual system, including screens and image generators; an audio system; and a motion system that gives the driver vehicle vibration and motion linked with the driver's operations.

HMI (human–machine interface/interaction) Interface or interaction between users and computer-based systems. The systems provide the users with visual, auditory, and/or haptic information. The users operate the systems using input devices, including a remote controller, a touch panel, and their voice.

Human-centered design Interface design of information contents or information devices that is adaptive to human cognitive and/or physical functions. In the automotive technology, research on the driver characteristics in perceptual, cognitive, and operational functions while driving is conducted for the purpose of applying the research findings to the interface design development of driver assistance systems, such as car navigation systems.

Instrumented vehicle A passenger vehicle equipped with various sensing technologies to detect and track internal and external conditions and a driving recorder system to save the measured data.

ITS (intelligent transport systems) Application of information, communication, and sensor technologies to multiple modes of transportation, including road, rail, air, and waterborne transports. Expected benefits by the introduction of ITS are reduction of traffic accidents, mitigation of traffic congestion, environmental improvement, positive economic impact, etc.

Naturalistic driving behavior Observation of driving behavior that takes place in its natural setting. The drivers are given no special instructions, no experimenter is present, and the data collection instrumentation is unobtrusive.

Preparatory behavior Driving behavior that occurs before making a turn at an intersection. This behavior includes activation of turn signal, release of the accelerator pedal, movement of driver's right foot to cover the brake pedal, and onset of pressure on the brake pedal.

Definition of the Subject and Its Importance

Vehicle navigation and communication systems play a role of an information center in road-traffic environments and a key component for realizing ITS (intelligent transport systems). The most popular function of the in-vehicle navigation systems is route guidance. The route guidance system presents drivers real-time, step-by-step driving instructions, such as preparation for turns, exits, or road changes. This function helps drivers choose and maintain efficient routes with less mental effort.

However, an inadequate interface design may lead to driver distraction or inattention to the primary driving tasks. This results in a reduction of driver acceptance of the route guidance as well as a reduction of driving safety. It is important to develop interface design of route guidance system in order for drivers to recognize the displayed contents correctly.

The in-vehicle navigation systems provide drivers with visual and auditory route guidance instruction as

well as a digital map with road information. The auditory message of the route guidance is usually presented three times while approaching the target intersection [1]. First, about 700 m before the target intersection, auditory information is provided for drivers to select the lane appropriate for the next turning direction. Second, guidance about 300 m before the intersection is presented when drivers can recognize a landmark. Third, just prior to reaching the intersection, a voice message (e.g., "turn to the right soon!") attracts attention to make the turn and can trigger a change in the driver's maneuvers from moving straightforward to preparing to make the turn. This essay focuses on the presentation timing of the last voice guidance. This last instruction is very important for the maneuver of tracing the provided route, especially where drivers have difficulty in recognizing the landmarks around the turning point due to complex road-traffic environments. Too early or too late presentation timing of the last message may lead to drivers' annoyance with the provided guidance and to drivers' navigational error, such as passing over the turning point. It is important to develop the presentation timing of the last voice guidance that enables drivers to begin preparing to make a turn at the same location as in their typical driving pattern. Understanding of the typical onset location of driver preparatory behavior for making a driver's side or curb side turn is necessary in the development of the optimal presentation timing.

When drivers missed correct turn, the navigation systems search the route again and update the guidance information. Rerouting requires a little time, and this delay may lead to a reduction in driver utilization of the route guidance. Prediction of the onset location of the driver's preparatory maneuvers prior to reaching the target intersection can contribute to enhancing navigation rerouting functions. For example, if a driver did not begin to prepare to make a turn after reaching the predicted onset point, the in-vehicle system can assess the driver's navigational error, determining whether he/she made an error by identifying an incorrect turning point or did not notice the information provided. The system can then reissue visual and/or auditory assistance information about the location of the target intersection or search the alternative route soon and present the updated route guidance. Modeling of the typical preparation behavior based on the driving behavior data is necessary to predict the onset location where the driver begins to prepare to make a turn under a specific road-traffic situation.

This essay describes a methodology of understanding drivers' typical preparatory behavior in a real road-traffic environment and of constructing preparatory behavior model using the behavioral data measured on the real road.

Introduction

Driving tasks have three levels of hierarchical structure: strategic, tactical, and operational [2]. The strategic level involves trip planning, determination of trip destination, and route choice, as well as general considerations about driving, including an evaluation of the costs, risks, satisfaction, and comfort. Tactical behaviors involve decision-making, such as the driving speeds, headway distances, and gaps in traffic. Operational behaviors involve the stable driving of the vehicle, including acceptable steering, and moderate acceleration and braking. The behavior at higher level has an effect on that at lower level, e.g., maneuvers at the tactical level (turning, overtaking, obstacle avoidance, and gap acceptance) must meet the criteria derived from the general goals set at the strategic level, and vice versa: e.g., a driver selects a route that includes many roads with less traffic volume when the driver likes a large margin to vehicles in the vicinity of his/her car.

Route guidance of in-vehicle navigation systems supports drivers' trip planning at the highest level of the driving tasks. In Japan, the route guidance systems presented a digital map, vehicle's current location on the map, and visual and auditory route guidance instructions to a destination as HMI in 1990s when the navigation systems began in widespread use. On the other hand, turn-by-turn displaying methods just indicating the next turning direction on the display were introduced in U.S. and Europe. Historically, the first type of the route guidance systems was developed by Honda in 1981. The contemporary navigation system indicating the vehicle's location on a map shown on a display device was mounted on the Crown of Toyota in 1987. Then, the first practical route guidance system which had real-time presentation of the vehicle location on a road map was the navigation system mounted

on the Nissan Cima, developed by Sumitomo Electric Industries, Ltd. in 1989. Please see literature [3] for development and commercialization process of navigation systems in Japan.

Route guidance systems using a digital map are now on the market all over the world. Recently, the navigation systems have more sophisticated functions, including real-time update of route guidance based on dynamic road-traffic information. These systems are integrated with several kinds of road sensors and road-traffic management center via multiple media. The systems provide the driver with the route guidance which takes into account current traffic jam conditions. In addition, the navigation systems present drivers with information on various facilities around the driving area, such as a car park.

The advanced functions of the in-vehicle navigation systems are expected to enhance road safety as well as to assist drivers in choosing and maintaining efficient routes with lower mental workload. However, the use of the new systems has strong concerns about driver confusion and distraction that can lead to an interference with the primary driving task and a reduction of driver acceptance of the provided information. Therefore, research on in-vehicle navigation systems with a human-centered design approach has been conducted and focused on the development of the interface design that is adapted to driver behavior in driving contexts [4].

The route guidance of in-vehicle navigation systems consists of visual and auditory information. The systems display a digital map with the route to a destination in real time while traveling. About 300 m before the target intersection, the systems present a magnified map of the vicinity of the intersection and the direction of the upcoming turn. The provision of the magnified map helps drivers identify the location of their next turn. Thus, visual information accompanying the magnified map is helpful in reducing drivers' misunderstanding of the target intersection. Human factor studies have investigated the information contents based on drivers' cognitive map and landmark designs that are suitable and acceptable to drivers [5–9].

In addition to the visual information, the navigation systems present auditory route guidance three times while approaching an intersection where to turn. The presentation timing is about 700 m before the intersection, about 300 m before it, and just prior to reaching it. The distance required for a lane change while approaching the target intersection, investigated by an empirical survey on urban roads [10], is "700 m." The distance where a landmark around the intersection (e.g., intersection sign, traffic signal, configuration of intersection) becomes recognizable and drivers identify the turning point is "300 m." The last guidance is important for drivers to leave the current road correctly because drivers frequently cannot recognize the landmarks around the turning point because of complex road-traffic environments, especially in urban areas. It is important to present the last guidance with appropriate timing that helps drivers begin preparing to make a turn at the same location as in their usual driving pattern. However, there may be some variations in the driving operations because the driving behavior is influenced by the road-traffic conditions [11]. Measurement and analysis of driver preparatory behavior prior to turning in an actual road environment are required and should clarify the influence of various road-traffic conditions – such as the presence of vehicles to the front and rear, the structure of the intersection, and the condition of the road surface – on the driver preparatory maneuvers. This essay focuses on the presentation timing of the last route guidance before making a driver's side turn at an intersection and deals with the existence of a lead and following vehicle and the intersection type (standard intersection, T-junction, and intersection after a curve). The behavior data described in this essay were collected on roads in Japan, where driving is on the left side of the road. Therefore, the right turn as driver's side in this essay corresponds to a left turn where driving is on the right side of the road, e.g., the United States.

Current navigation systems reroute when a driver failed to identify the turning point or failed to notice the provided guidance and he/she passed over the target intersection. Prediction of the onset location of the driver preparatory behavior can contribute to detecting the driver's navigational error before passing through the intersection at which to turn. The in-vehicle system can assess driver's failure of identifying the required turning point or his/her unawareness of the information provided if the driver does not begin to prepare for making a turn after reaching the predicted location.

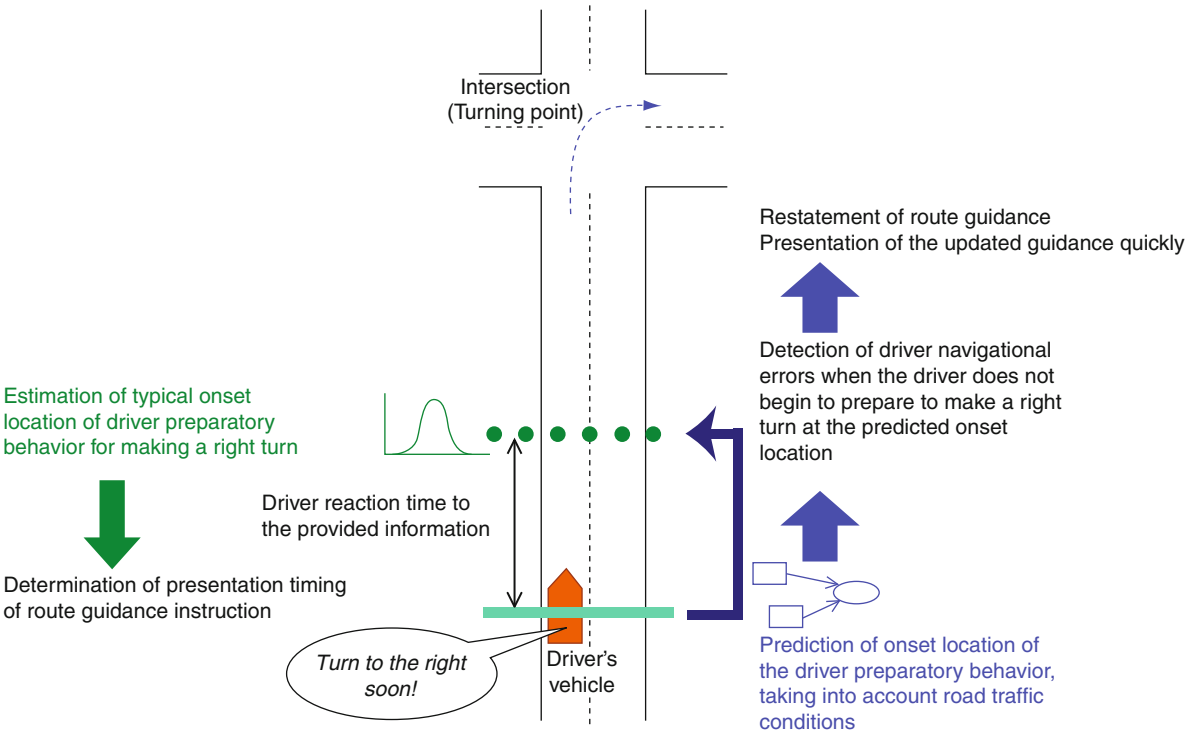
Thus, the preparatory behavior prediction can lead to further advanced functions of the route guidance systems, such as restating the route guidance and quick rerouting and providing the updated route. The former function is expected to reduce the driver's error of passing over the target intersection, and the latter is expected to enhance usability of the route guidance systems. The influence of the road-traffic environments on the driver preparatory behavior should be evaluated quantitatively in order to predict the onset of the behavioral event. To accomplish this, a driving behavior model should be constructed based on naturalistic behavioral data measured repeatedly in a real environment.

Figure 1 presents a design concept of the presentation timing adapted to driver's typical preparatory behavior and the detection of driver navigational error based on prediction of the preparatory behavior onset. This essay describes a method of estimating typical onset location of driver behavior for preparing to make a driver's side turn and of predicting the onset location using statistical driver model.

Behavioral Data Collection in an Actual Road Environment

Overview of Data Collection Methods

Driving a vehicle in traffic is a task that is influenced by road-traffic environments. The preparatory behavior before making a turn is also influenced by traffic conditions while approaching an intersection where to turn and by road environments of the target intersection. It is essential to collect data on the external factors that affect the driver preparatory behavior in addition



Driver Behavior at Intersections. Figure 1
Design concept of the presentation timing of last voice guidance compatible with driver behavior and the detection of driver navigational error. The typical onset location of driver preparation before making a driver's side turn is estimated based on measurement of naturalistic behavior data while approaching a target intersection. The driver navigational error is detected based on comparison between the actual behavior and prediction of the onset location using statistical driver model

to the data on the driver's operations for the preparations for making a turn.

A driver experiences only a few kinds of traffic conditions for a short term. For example, 1 day, there is a lead vehicle driving slowly in front of the driver's vehicle on one leg of a route; another day, there is a lead vehicle driving faster on the same leg; and the other days, he/she has no preceding vehicles on the same leg. Thus, a few days are too short to understand the driver's typical preparatory behavior under real traffic conditions. It is necessary to carry out repeated experiments on the same roads over a long term. The long term for the data collection is effective in recording the driver preparatory behavior on various kinds of road conditions, including dry, wet, and rainy roads. Analysis of the driving behavior that is measured on different intersections with various structures is necessary to investigate the influence of the road structures on the driver preparatory behavior. A driving route, which includes several kinds of intersection types – e.g., intersection on a straight road, intersection at the end of the road (T-junction), intersection after a curved road, and intersection after a slope – should be taken into account when designing and conducting an experiment.

Methodologies for the behavior data collection are point measurements, driving simulator experiments, and field experiments. Traffic-flow point measurements are conducted in which the number of vehicles, the time intervals between passing vehicles, and the traveling speeds are recorded via video cameras fitted at the site [12, 13]. However, the point measurements cannot clarify onset location and timing of driver decelerating operations.

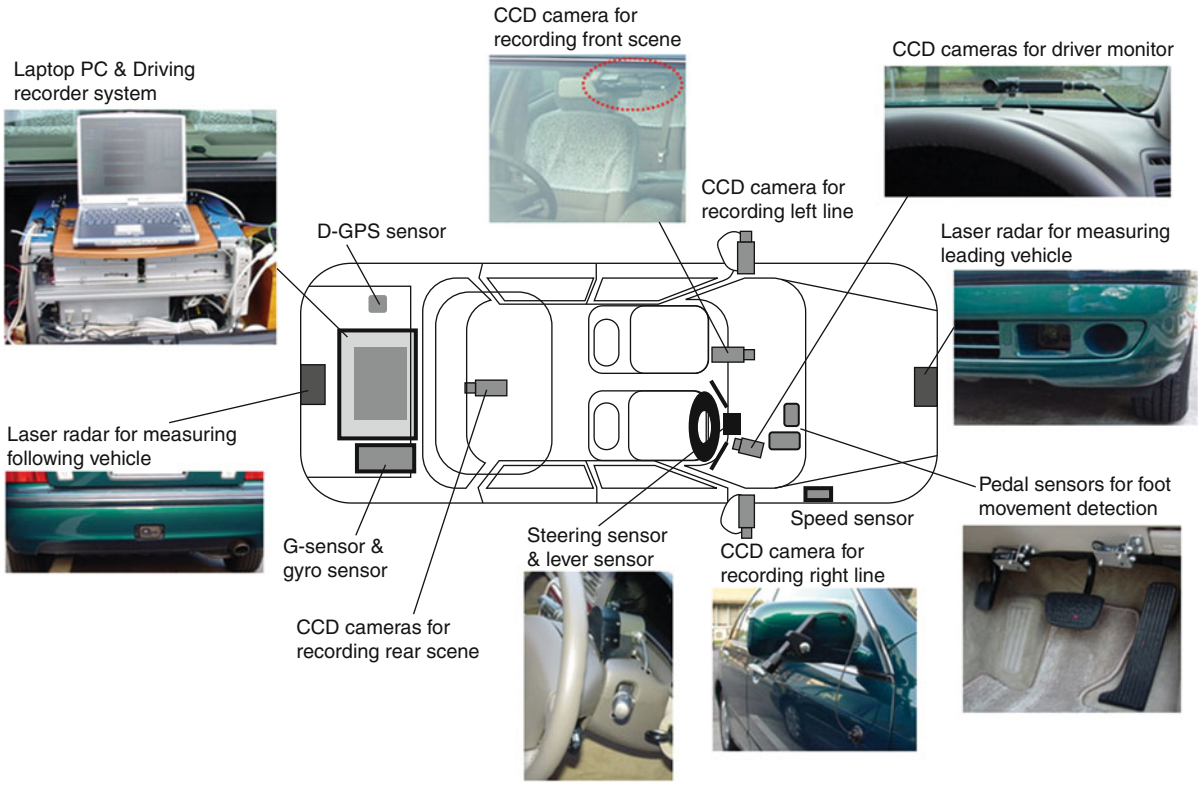
Driving simulators have been used as experimental tools to measure driver operations and to examine the relationships between the external factors and the driving behavior [14–16]. Recently, image generation technologies, such as construction of geometric road structures based on CAD data and construction of scenery along the roads using texture mapping, have been applied to the development of realistic road environments [17]. However, the simulator cannot fully reproduce natural driving behavior in a real road-traffic environment due to mechanical restrictions, e.g., the lack of longitudinal movement during deceleration and the lack of speed and distance perception [18, 19]. These limitations negatively influence the driver operations before making a driver side turn. For example, a driver tends to apply the brake pedal several times before coming to a full stop at the target intersection in the driving simulator environment, whereas the stroke of the brake pedal is executed once and the driver presses down on the brake pedal gradually in the real road environment. Therefore, the onset location and timing of the preparatory behavior at a real intersection differ from those at the intersection of a driving simulator where the geometric structure and roadside landmarks are represented exactly similar to the real intersection.

Although repeated field experiments for a long period have a risk for traffic accidents, the real-world experiments are the most effective in collecting naturalistic data on the driver preparatory behavior. The recent development of memory devices to record large-scale data and sensing technologies to detect and track external conditions contributes to the realization of the field data collection. The following two approaches are used for the repeated experiments on an actual road: (1) installation of memory devices and sensors into drivers' owned passenger vehicles and (2) development of instrumented vehicles equipped with various sensors and a driving recorder system.

The representative project of the former method is the 100-Car Naturalistic Driving Study by Virginia Tech Transportation Institute [20]. In this project, several data collection devices, including speed sensor, GPS, accelerometers, front and rear radar sensors, and five channels of digital video, are installed into participants' owned vehicle. The large-scale behavior data have been recorded over 13 months with the primary purpose of collecting precrash and near-crash naturalistic driving data. The advantages of using privately owned vehicles are to eliminate the participants' unfamiliarity with an experimental vehicle and to eliminate their awareness of being monitored. This method has limitations from the viewpoint of driver preparatory behavior analysis in extracting a large amount of behavior data around the target intersections from the large-scale data.

Use of Instrumented Vehicles to Collect Behavioral Data

Instrumented vehicle is developed to collect driver behavior data as well as to measure the vehicle status and the



Driver Behavior at Intersections. Figure 2
Overview of the AIST instrumented vehicle. Several sensors and a driving recorder system are fixed inside the trunk where the drivers cannot see the instruments, thus encouraging naturalistic driving behavior. A participant drives alone this vehicle during measurement trials

traffic conditions around the vehicle. Figure 2 presents the AIST instrumented vehicle [21]. The following data sets are collected using the instrumented vehicle:

- Driving speed
- Vehicle acceleration (longitudinal, lateral, and vertical)
- Angular velocity (roll, pitch, and yaw)
- Geographical position of the vehicle
- Relative distance and speed to leading vehicle
- Relative distance and speed to following vehicle
- Application of accelerator and brake pedals
- Position of driver's right foot (covering the pedal without pressing)
- Steering wheel angle
- Turn signal activation
- Visual images (forward and backward traffic scenes, left and right lane positions, and the driver's face)

A speed sensor detects the driving speed using a speed pulse signal. A G-sensor detects the vehicle acceleration in longitudinal, lateral, and vertical directions, and a gyro sensor measures the angular velocity in roll, pitch, and yaw directions. A D-GPS sensor obtains the geographical position of the vehicle. Laser radar units fixed within the front and rear bumpers record the relative distance and relative speed to the leading and following vehicles. Potentiometers measure the applications of the accelerator and brake pedals. Laser sensors fitted above the pedal surfaces detect the position of the driver's right foot (covering the accelerator or brake pedal without pressing). Encoders added to the steering wheel and turn signal lever record the steering wheel angle and turn signal activation. A total of five CCD cameras are used to record visual images, including the forward and backward traffic scenes, left and right lane positions, and the

driver's face. These data are recorded on a laptop computer and mobile hard disks via a driving recorder system. The sampling frequency of all data except the visual images is 30 Hz; the visual images are sampled at eight frames/s. The abovementioned instruments are arranged so as to be as unobtrusive as possible (the recorder system is fixed inside the trunk of the vehicle) in order to encourage naturalistic driving behavior.

The disadvantage of using the instrumented vehicle is participants' unfamiliarity with the experiment vehicle. Practice drives before measurement trips are important in order for the participating drivers to be familiar with the instrumented vehicle and the experimental atmosphere as well as to drive from the start to the end without using any assistance, such as a map, a vehicle navigation system, and passengers. The driving route is predetermined, which included several left and right turns. The participant rides alone in the vehicle during the experiment trials, whereas the experimenter rides with the participants on the first day of the trials in order to confirm that the participants are following the correct route. Thus, the data from the first day are not used for subsequent data analyses and data modeling. The participating drivers are instructed only to drive in their typical manner.

Analysis of Data Distribution Based on Traffic and Road Environments

Driver preparatory behavior before making a turn is composed of turn signal activation and deceleration. The decelerating operations involve releasing the accelerator pedal, moving the right foot from the accelerator pedal to the brake pedal, and pressing the brake pedal. In this section, data analysis of the turn signal use and the foot movement to cover the brake pedal is introduced as representative maneuvers of the driver preparatory behavior. The accelerator pedal is released not only to decelerate for making turns but also to maintain an appropriate driving speed [22]. The brake pedal application is an operation after driver's decision-making on leaving the current road, and it is influenced by driver factors (such as driving skill) other than the decision to leave the current road [23]. Only the data acquired when there was a green light at the target intersection are analyzed in the following sections. The influence of traffic conditions on driver preparatory behavior differs between encountering green and red traffic lights. It is recommended to exclude the behavioral data recorded at the red traffic light when developing the presentation criteria of route guidance based on the measurement of the typical preparatory behavior [24].

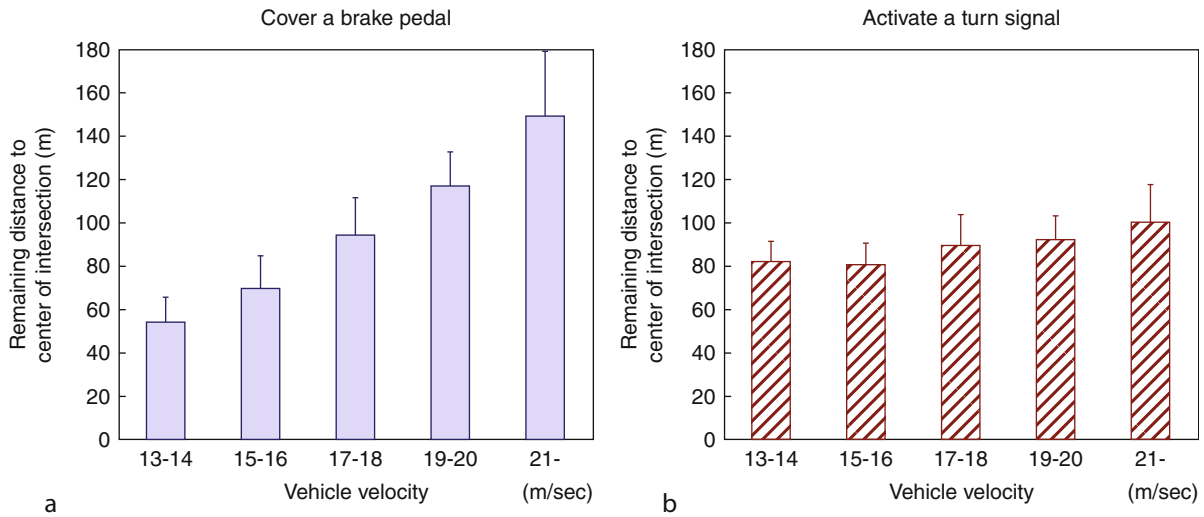
Influence of Vehicle Velocity on Driver Preparatory Behavior

Vehicle velocity while approaching an intersection influences the driver decelerating operation before making a driver's side turn. In contrast, the onset location of the turn signal activation is almost stable regardless of the vehicle velocity. Figure 3 presents the onset locations of driver preparatory behavior in the five categorized vehicle velocities. Field experiments using the AIST instrumented vehicles were conducted, and the preparatory behavior data were collected before turning at an intersection with two traffic lanes and a designated lane for making a driver's side turn. The number of participants was four (three males and one female). The average age of the participants was 34.8 years, and the average driving experience was 16.3 years. The total trips for each participant were 40 over a period of about 2 months (weekdays). The four participants started driving on the identical route (total mileage, 15 km) at 10-min intervals. Please see literature [24] for detailed experiment methods and analyses. The vehicle velocity is measured when the driver moves the right foot from the accelerator pedal to the brake pedal because the velocity, when covering the brake pedal and when activating the turn signal, is almost the same.

The remaining distances to the center of the intersection when covering the brake pedal are directly proportional to the driving speeds, and the onset locations of the foot movement become closer to the intersection when approaching it more slowly. In contrast, the proportional relation between the speed and the turn signal use is not remarkable. The onset locations of the turn signal operation do not change in any categorized driving speeds.

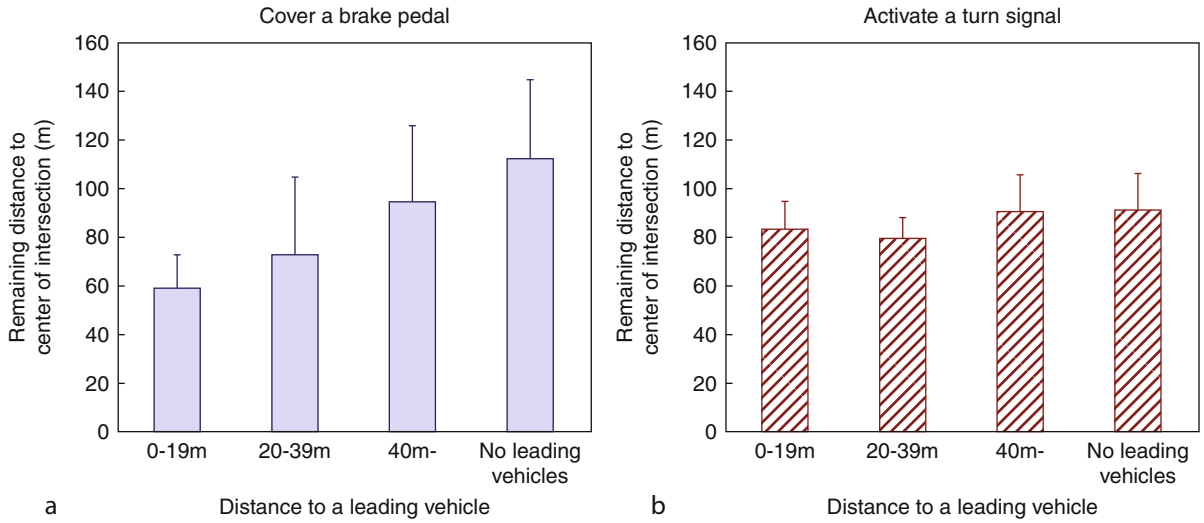
Influence of Traffic Conditions on Driver Preparatory Behavior

Figure 4 presents the onset location of driver preparatory behavior in the four categories of the relative



Driver Behavior at Intersections. Figure 3

Relationship between the onset location of driver preparatory behavior and vehicle velocity. The graphs present averages and standard deviations of the remaining distances to the center of the intersection when each behavioral event occurs in each category of the vehicle velocity. The data were collected at an intersection with two traffic lanes and a designated turn lane



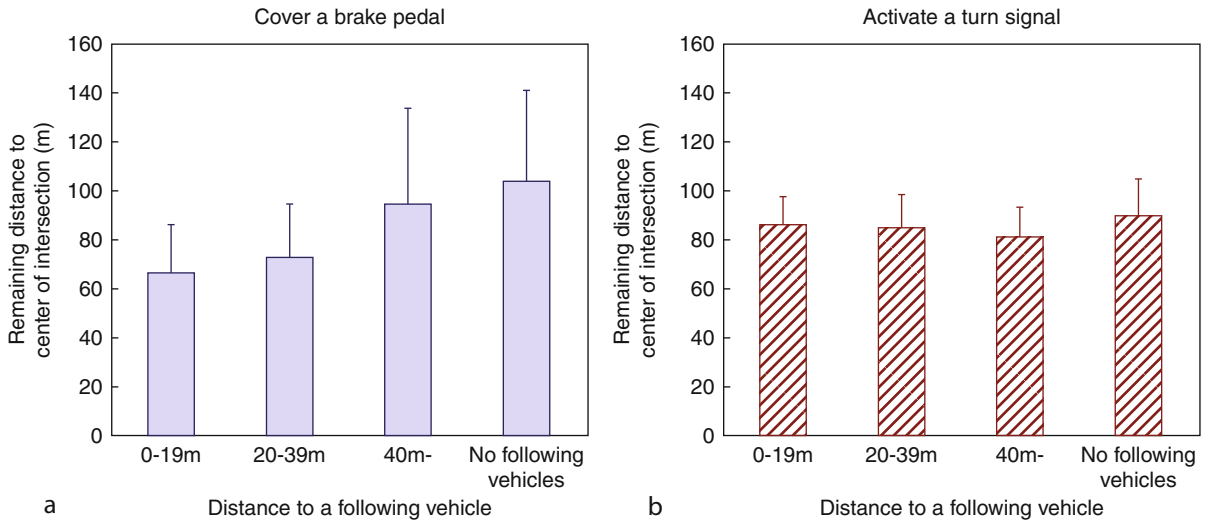
Driver Behavior at Intersections. Figure 4

Relationship between the onset location of driver preparatory behavior and relative distance to a leading vehicle. The graphs present averages and standard deviations of the remaining distances to the center of the intersection when each behavioral event occurs in each category of the relative position to the leading vehicle. The data were collected at the intersection with two traffic lanes and a designated turn lane

distance to the lead vehicle, including the absence of a leading vehicle. The data used are the same as those in Fig. 3. The relative distances from the leading vehicle are measured before the driver enters the designated turn lane. Here, the leading vehicle is a forward vehicle that travels straight toward the target intersection. The remaining distances to the center of the intersection when covering the brake pedal while driving with a leading vehicle are shorter than those during drives without a leading vehicle. In addition, shorter relative distances to the leading vehicle lead to shorter remaining distances at the onset of covering the brake pedal. On the other hand, the onset location of turn signal activation is at a remaining distance of between 80 and 90 m to the center of the target intersection. The turn signal use is independent of the existence of and relative distance to a leading vehicle.

Figure 5 presents the onset location of driver preparatory behavior in the four categories of the relative distance from a following vehicle, including the absence of a following vehicle. These data are also the same as those in Fig. 3. The relative distances from the following vehicle are measured before the driver enters the designated turn lane. Here, the following vehicle is a vehicle that follows the driver's vehicle in the same driving lane. The remaining distances when covering the brake pedal while driving with a following vehicle are shorter compared to driving without a following vehicle, particularly for a shorter relative distance between vehicles. On the other hand, the remaining distances when the turn signal is activated are almost stable in the four categories of the relative distance from the following vehicle, ranging from 80 to 90 m.

The vehicle velocity while approaching the intersection is correlated with the relative distance between the driver's vehicle and a leading or following vehicle. It is hypothesized that only the slower driving speed leads to a shorter remaining distance when covering the brake pedal. At a specific range of vehicle velocity (from 16.0 to 17.5 m/s), the onset locations when covering the brake pedal were compared among the four situations: driving with both leading and following vehicles, driving with only a leading vehicle, driving with only a following vehicle, and driving without either a leading or following vehicle. The remaining distance at the onset of covering the brake pedal was the shortest when driving with both leading and



Driver Behavior at Intersections. Figure 5

Relationship between the onset location of driver preparatory behavior and relative distance to a following vehicle. The graphs present averages and standard deviations of the remaining distances to the center of the intersection when each behavioral event occurs in each category of the relative position from the following vehicle. The data were collected at the intersection with two traffic lanes and a designated turn lane

following vehicles. The onset location of covering the brake pedal was closer when driving with only a leading vehicle than that when driving with only a following vehicle.

Drivers begin to decelerate at a point closer to the center of the intersection under car-following conditions compared to free-running conditions at the same driving speed. The preparation point is the closest to the target intersection when drivers prepare to make a turn with a following vehicle in addition to a leading vehicle. Drivers following a forward vehicle pay attention to that vehicle in order to respond quickly to its actions [25, 26]. Additionally, they pay attention both forward and backward in so-called platoon conditions under which they travel a short distance to leading and following vehicles. Intense concentration on vehicles surrounding the driver's vehicle may lead to a delay in the onset of driver decelerating maneuvers before making a driver's side turn at an intersection.

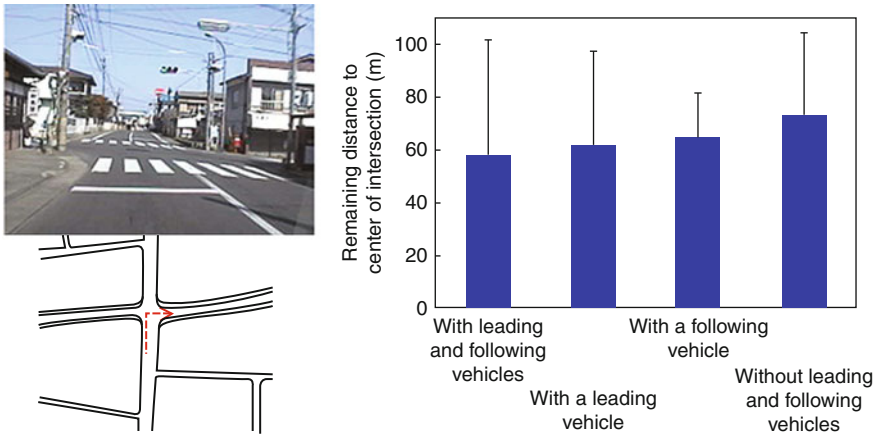
The analyses of turn signal use within the specific range of the vehicle velocity reveal the fixed location of turn signal activation regardless of the traffic conditions surrounding the driver's vehicle. As mentioned from Figs. 3–5, the remaining distances to the center of the target intersection when activating the turn signal are almost constant in all ranges of the vehicle velocity and of the relative distance from the leading or following vehicle. The analyzed intersection has a designated turn lane (the distance between the entrance of the designated lane and the center of the intersection is 70 m). This road alignment feature may contribute to the constant location at which the drivers activate the turn signal.

Influence of Road Structure on Driver Preparatory Behavior

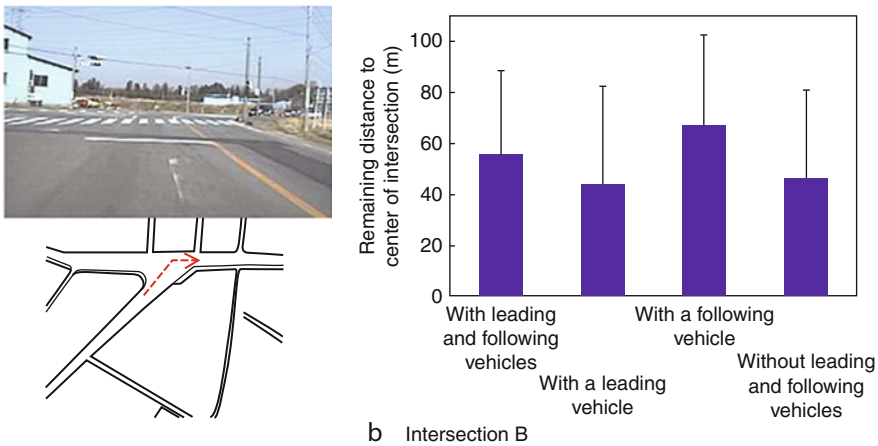
Traffic conditions (the existence of leading and/or following vehicles) influence the driver decelerating maneuver at a specific intersection with two traffic lanes and a designated turn lane. Moreover, the road structures of intersection where to turn influence the relations between the traffic conditions and the driver preparatory behavior for decelerating before a driver's side turn. Figure 6 presents the onset locations of covering the brake pedal in the four categorized traffic conditions at three kinds of intersections. The data collection methods were as follows: number of participants, eight (five males and three females); average age of the participants, 38.1 years; average driving experience, 18.1 years; total trips for each participant, 40 over a period of about 2 months (weekdays). Four participants started driving on the route, including the target intersections, at 10-min intervals. The other drivers made recorded trips at another period following the same procedures. Please see literature [27] for detailed experiment methods and analyses. Intersection A is a standard right-angle intersection on a straight road with one traffic lane. Intersection B is a T-junction at the end of the road. Intersection C is also a T-junction, and this intersection exists after drivers drive along a curved road (curve radius, 126.9 m).

At intersection A, the remaining distances while driving with leading and following vehicles and while driving with only a leading vehicle are shorter than those while driving without a leading or following vehicle. At intersection B, the onset locations of covering the brake pedal tend to be longer to the center of the intersection when there is a following vehicle compared to only a leading vehicle and without a leading or following vehicle. At intersection C, the remaining distances to the center of the intersection when covering the brake pedal are longer when there is a vehicle in front. The onset location is the longest among the four traffic conditions when driving with both leading and following vehicles.

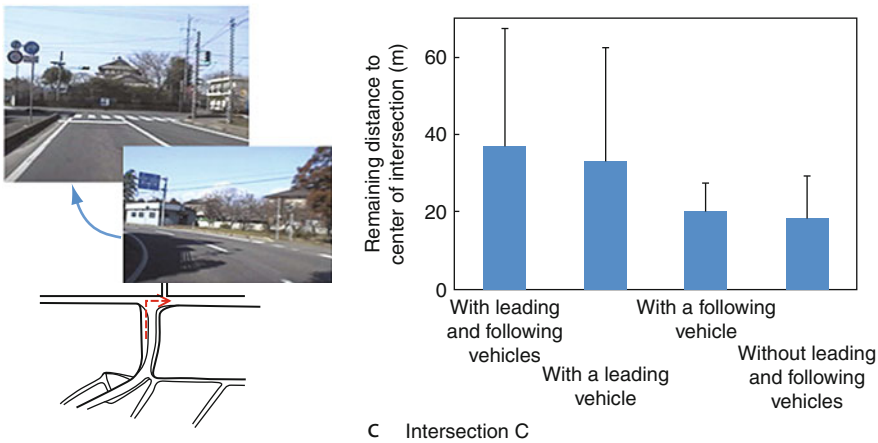
The comparison between the standard intersection and the intersection with a designated turn lane indicates that the number of traffic lanes has less influence on the relationship between the traffic conditions and the driver decelerating behavior before making a driver's side turn. The behavioral analysis at the intersection after a curved road suggests that drivers begin to prepare for deceleration at an earlier point while approaching the intersection with leading and following vehicles and with only a leading vehicle compared to without such vehicles. Drivers cannot see the landmarks, including traffic lights, and cannot recognize the remaining distance to the turning point until just before the intersection after a curve. At the standard intersection, drivers can see and recognize the location of the intersection, and they may adapt their behavior to the leading vehicle until they are closer to the intersection. In contrast, drivers change their



a Intersection A



b Intersection B



c Intersection C

Driver Behavior at Intersections. Figure 6

Relationship between the onset locations of driver preparatory behavior and traffic conditions at three kinds of intersections. The graphs present averages and standard deviations of the remaining distances to the center of the intersection when the drivers move their right foot from the accelerator pedal to the brake pedal in each category of the traffic conditions

driving mode from following the leading vehicle to preparing to make a driver's side turn at an earlier point when the intersection cannot be recognized in advance.

Drivers begin to decelerate farther from the T-junction on a straight road when they have a vehicle following. In this case, drivers may intend to use their brake lights as a signal to the following driver to avoid a rear-end collision by the following vehicle. At T-junction, the leading vehicle always decelerates and turns right or left. Drivers can anticipate the deceleration of the lead vehicle, and they may pay more attention to the following vehicle compared to the movement of the lead vehicle while approaching the T-junction with leading and following vehicles. A driver's recognition of a following vehicle may lead to earlier preparation for deceleration. Intersection C is also a T-junction, but the onset location of covering the brake pedal when driving with only a following vehicle is similar to driving without leading or following vehicles. Drivers should pay attention to the inside of the curve during curve negotiation [28]. Thus, drivers might not recognize the presence of a following vehicle while approaching an intersection after a curve. When drivers notice the following vehicle before entering the curve, they may allocate more resources to the curve negotiation and may not pay attention to the rear vehicle.

Recommendations for Presentation Timing of Route Guidance

The presentation timing of route guidance instruction compatible with driver's typical preparatory behavior is recommended to enhance driver acceptance of the provided information. Data analyses described in section "Influence of Vehicle Velocity on Driver Preparatory Behavior, Influence of Traffic Conditions on Driver Preparatory Behavior, Influence of Road Structure on Driver Preparatory Behavior" suggest that it is effective to change the timing of the last route guidance presentation based on the vehicle velocity, the existence of and relative distance to leading and following vehicles, and the road structures of the turning point.

First of all, the presentation timing should be changed corresponding to the driving speeds while approaching a target intersection. Driving faster requires earlier presentation timing compared to driving slower. In addition to the vehicle velocity, the traffic conditions surrounding the driver's vehicle should be taken into account. Presenting the instruction closer to the intersection would be effective for guidance and compatible with driver preparatory behavior while approaching right-angle intersections on straight roads under close car-following conditions, whereas earlier presentation would be acceptable while driving with a leading and/or following vehicle at long range and while driving without leading or following vehicles. Moreover, the route guidance presentation with different timings at intersections with T-junctions and at intersections after curves is recommended. Earlier guidance would effectively enhance drivers' utilization of the instruction while approaching a T-junction on a straight road when driving with a following vehicle, independent of the existence of the lead vehicle. At intersections after curves, drivers cannot see or recognize the relation between their vehicle and the turning point until just before the target intersection. In these cases, the route guidance instructions should be provided earlier in car-following situations to improve the drivers' acceptance of the presented information.

Modeling of Driver's Preparatory Behavior Before Making Driver's Side Turn

Preparatory Behavior Model Using Structural Equation Modeling

Data analyses of driver preparatory behavior before making a driver's side turn at a specific intersection with a designated turn lane reveal linear relationships among the vehicle velocity, the traffic conditions in the vicinity of the driver's vehicle, and the preparatory behavior (sections "Influence of Vehicle Velocity on Driver Preparatory Behavior and Influence of Traffic Conditions on Driver Preparatory Behavior"). The onset locations of the foot movement to cover the brake pedal are positively correlated with the driving speeds or the remaining distances to leading and following vehicles. In contrast, the onset locations of the turn signal operation are almost constant, independent of the speeds and the traffic conditions. Structural equation modeling is an efficient modeling technique for quantifying the impact of independent variables on the dependent variables that suggests linear

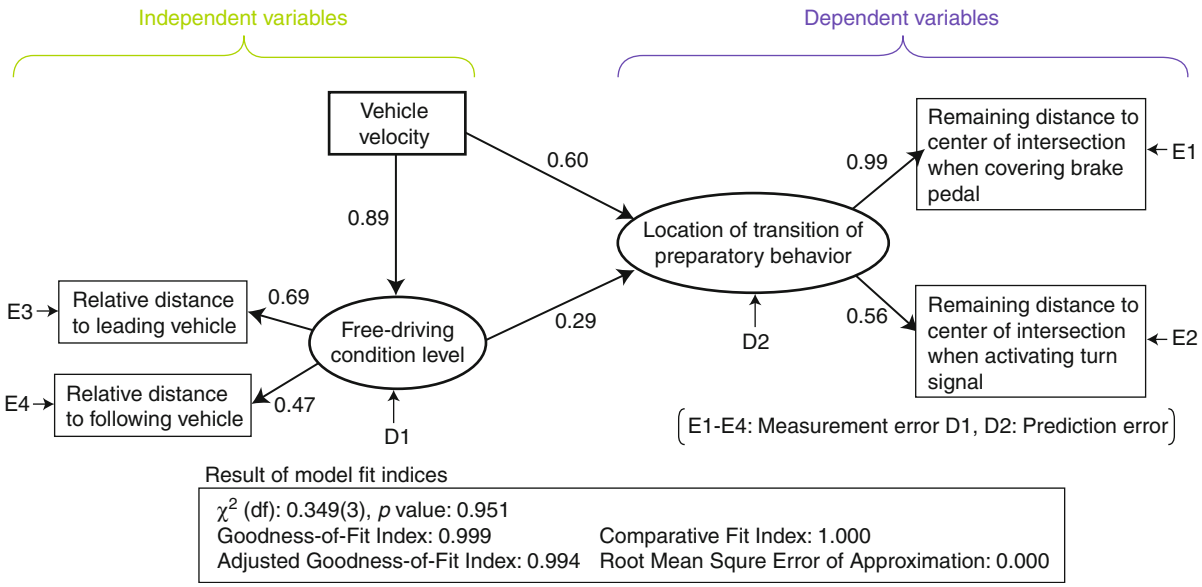
relations [29]. Thus, the structural equation model can be applied to describe quantitatively the relationships among the vehicle velocity, the relative distances to the leading and following vehicles, and the onset locations of covering the brake pedal and activating the turn signal.

Structural equation modeling represents the interrelations among a set of variables comprising a model hypothesized by theory and empirical research. Observed and latent variables are used in the structural equation model: observed variables are directly collected and measured, and latent variables are inferred from the set of observed variables. The parameters specified in the structural equation model are estimated so as to minimize differences in the variance–covariance matrices between the hypothesized theoretical model and the measurement data. Various kinds of model fit indices evaluate whether or not the covariance matrix of the hypothesized model is as close as possible to that of the observed variables [30].

In the structural equation model of driver preparatory behavior, the following five observed variables are used: the driving speed, the relative distance to the leading vehicles, the relative distance to the following vehicles, the remaining distance to the center of the target intersection when covering the brake pedal, and the remaining distance to the center of the target intersection when activating the turn signal. Subsequently, the constructed model is applied to a prediction of the onset location of the driver preparatory maneuvers. Therefore, the independent variables, the driving speed and the headway and rear distances, at the onset of releasing the accelerator pedal are used in the model construction. Here, the release of the accelerator pedal is defined as the final release maneuver of the accelerator pedal, leading to right foot movement to cover the brake pedal. The cases where there are no leading and/or following vehicles when releasing the accelerator pedal have missing data for the relative distances to the leading and/or following vehicles. To overcome this limitation, having no leading and/or following vehicles around the driver's vehicle is defined as the condition in which leading and/or following vehicles are such a long way from the driver's vehicle that the laser radar cannot detect and measure them. The relative distances to the leading and following vehicles in the case of no leading and/or following vehicles are set to 100 m, a value which is beyond the detection limits of the laser radar unit.

The latent variables are important to obtain an acceptable model to data fit in the structural equation model specification of driver preparatory behavior. This is because the velocity and the headway and rear distances influence, not independently but interactively, the onset locations of driver preparations for making a driver's side turn. A latent variable, called the free-driving condition level, is input to the model specification as a latent factor describing the interaction between the leading or following vehicles and the driver's vehicle. This latent variable uses two indicator variables: the relative distance to the leading vehicle and the relative distance to the following vehicle. A higher free-driving condition level denotes that a driver drives without leading and following vehicles or remains a long distance from the front and rear vehicles. Conversely, a lower free-driving condition level denotes that a driver drives under close car-following conditions. In addition, a latent variable, called the location of transition to preparatory behavior, is introduced as a latent factor related to the onset location of the change in driving behavior from straight (going forward along the roadway) mode to the preparation mode. This latent variable comprises two indicator variables: the remaining distance to the center of the intersection when covering the brake pedal and the remaining distance to the center of the intersection when activating the turn signal. The location of transition to preparatory behavior denotes the degree to which a driver begins to prepare for making a driver's side turn at a long distance from the center of the intersection. A closer location of transition to preparatory behavior denotes that the onset location of the behavioral event is closer to the target intersection. Please see literature [31] for the details of the model construction.

Figure 7 presents a path diagram of the proposed structural equation model, the estimated path coefficients and factor loadings (standardized weights), and the results of model fit indices. The results of the model fit indices suggest that the estimated structural equation model fits the observed data well. The vehicle velocity and the free-driving condition level are positively related to the location of transition to preparatory behavior. The relation between the free-driving



Driver Behavior at Intersections. Figure 7

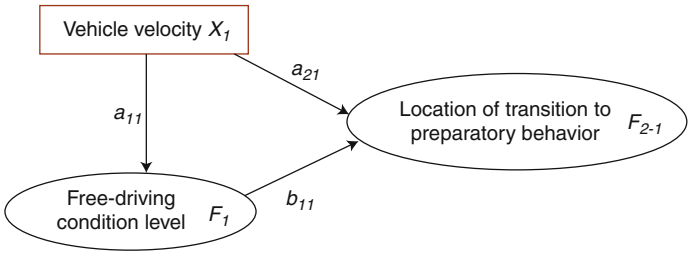
Structural equation model of driver preparatory behavior. The values in the path diagram present the standardized path coefficient (among vehicle velocity, free-driving condition level, and location of transition to preparatory behavior) and factor loadings (between the latent variables and each indicator variable). The model was estimated using the maximum likelihood method. Total data sets are 119 for 4 drivers

condition level and each indicator variable (relative distances to leading and following vehicles) is also positive, suggesting that the location of transition to preparatory behavior is directly proportional to the headway and rear distances. The factor loading from the location of transition to preparatory behavior to the remaining distance at the onset of covering the brake pedal is higher than the factor loading for the remaining distance when activating turn signal. This means that higher driving speeds and free-driving conditions contribute to an earlier onset location when covering the brake pedal and that the independent variables have less effect on turn signal activation.

At T-junctions and intersections after curves, the relationships between the traffic conditions and the onset location of driver decelerating maneuver are different from those at the right-angle intersections on straight roads. In these cases, the structural equation model should be estimated at each intersection. The path diagram is the same among the intersections, and the path coefficients and/or the factor loadings will be changed according to the relationships observed at each intersection. At T-junctions, the factor loading between the free-driving condition level and the relative distance to the following vehicle will be negative. At intersections after curves, the path coefficients from the vehicle velocity and the free-driving condition level to the location of transition to preparatory behavior will be negative.

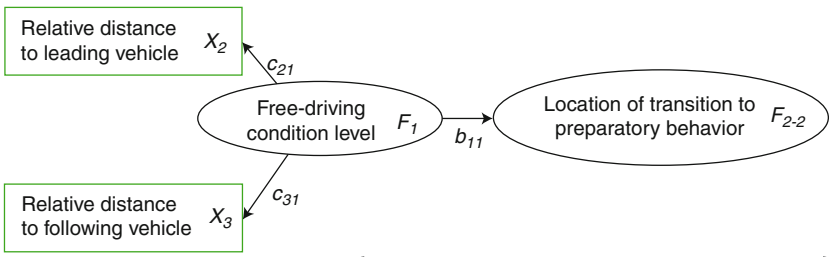
Prediction of Onset Location of Driver Preparatory Behavior

The path coefficients and factor loadings estimated in the structural equation model are used to predict the onset locations of driver preparatory operations. Figure 8 presents the method for predicting the onset locations of the driver preparatory behavior. The influences of the vehicle velocity on the location of transition to preparatory behavior are divided into direct influence and indirect influence via the free-driving condition level (Fig. 8a). The factor score for the free-driving condition level is calculated by using the factor loadings for the two indicator variables because the relationship between the free-driving condition level and the relative distances to the leading and



$$F_{2-1} = (a_{21} + a_{11} \times b_{11}) \times X_1$$

a Direct and indirect influence of vehicle velocity

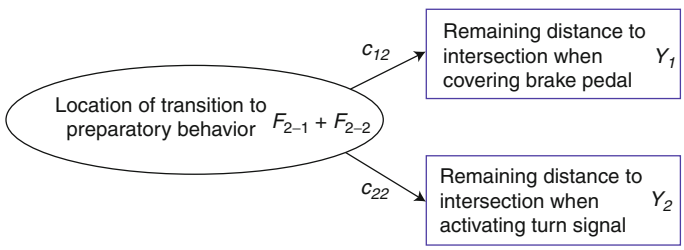


$F_1 = XR^{-1}C$

X is a matrix of standard scores of X_2 and X_3
R is a correlation matrix of X
C is a matrix of the factor loadings of c_{21} and c_{31}

$$F_{2-2} = b_{11} \times F_1$$

b Influence of relative distances to leading and following vehicles using factor score estimation



$$Y_1 = c_{12} \times (F_{2-1} + F_{2-2})$$

$$Y_2 = c_{22} \times (F_{2-1} + F_{2-2})$$

c Prediction of onset locations of foot movement to cover brake pedal and turn signal activation

Driver Behavior at Intersections. Figure 8

Procedures for predicting the onset locations of driver preparatory maneuvers. The procedures are divided into three steps: the effect of vehicle velocity on location of transition to preparatory behavior, the effect of headway and rear distances on location of transition to preparatory behavior, and the calculation of remaining distances to intersection when covering brake pedal and activating turn signal from the latent variable. (**a**) Direct and indirect influence of vehicle velocity. (**b**) Influence of relative distances to leading and following vehicles using factor score estimation. (**c**) Prediction of onset locations of foot movement to cover brake pedal and turn signal activation

following vehicles corresponds to a confirmatory factor model [32]. Then, this score is used to calculate the direct effect of the free-driving condition level on the location of transition to preparatory behavior (Fig. 8b). Finally, the remaining distances to the center of the intersection when covering the brake pedal and when activating the turn signal are predicted by using the factor loadings between the location of transition to preparatory behavior and the onset location of each behavioral event (Fig. 8c).

The onset locations of covering the brake pedal and activating the turn signal can be predicted by applying estimated path coefficients and factor loadings to the recorded vehicle velocity and relative distances to leading and following vehicles. The estimated structural equation model uses the driving speed and the headway and rear distances when drivers release the accelerator pedal. The onset location of releasing the accelerator pedal may be almost the same as the last route guidance presentation that is determined based on drivers' typical preparatory behavior. The route guidance system linked with laser radar units and a speed pulse signal source can predict the driver decelerating maneuver and turn signal operation when the last voice message is presented. Sometimes a driver does not begin preparation after reaching the predicted onset location. In such cases, the system can infer a driver error that he/she did not accurately identify the turning point due to a misunderstanding of the provided information or did not notice the information provision. The systems could then restate the voice route guidance or quickly reroute and present the updated route.

Future Directions

Further data analyses reveal that the relationships among the onset location of driver preparatory behavior, the traffic conditions, and the road environments before making a curb side turn are similar to those before making a driver's side turn described in this essay [27]. Thus, the findings of understanding and modeling of typical preparatory maneuvers for a driver's side turn can be applied to the route guidance presentation timing and restatement before a curb side turn.

Understanding the driver usual preparatory behavior and developing the structural equation model need a large amount of measured data. Recording of the large-scale data by one vehicle requires a few weeks. After a driver bought an in-vehicle navigation system and used it in his/her daily lives for a few weeks, the route guidance system becomes sophisticated: the presentation timing becomes compatible with his/her typical preparatory behavior, and the advanced functions of the restatement of route guidance and the quick update of route become available when he/she misses the turning point.

Data analysis and modeling approach can be applied to driving tasks other than preparations for turning at intersections. The linear relations between the driving behavior and the traffic conditions are found in the driver's side turn behavior, and the structural equation modeling is adopted. If the measured data sets suggest nonlinear relationships, probabilistic models, including a Markov dynamic model and a Bayesian network model, will be used for modeling and predicting the driving behavior in a real environment (e.g., [33, 34]).

Bibliography

Primary Literature

1. Kimura K, Marunaka K, Sugiura S (1997) Human factors considerations for automotive navigation systems – legibility, comprehension, and voice guidance. In: Noy YI (ed) Ergonomics and safety of intelligent driver interfaces. Lawrence Erlbaum Associate, Mahwah, pp 153–167
2. Michon JA (1985) A critical view of driver behavior models: What do we know, what should we do? In: Evans L, Schwin RC (eds) Human behavior and traffic safety. Plenum, New York, pp 485–520
3. Ikeda H, Kobayashi Y, Hirano K (2011) How car navigation systems have been put into practical use – development management and commercialization process-. Synthesiology – English Edition 3:280–289
4. Pauzie A (2005) Methodologies for driver-centred evaluation of perceptive and cognitive requirements of in-vehicle display. In: Proceedings of 11th international conference on human-computer interaction, Las Vegas, CD-ROM
5. Pauzie A, Daimon T, Bruyas MP (1997) How to design landmarks for guidance systems? In: Proceedings of 4th world congress on intelligent transport systems, Berlin, CD-ROM
6. Daimon T, Nishimura M, Kawashima H (2000) Study of driver's behavioral characteristics for designing interfaces of in-vehicle navigation systems based on national and regional factors. JSAE Rev 21:379–384

7. Kawai M, Kawasumi M, Nakano T, Yamamoto S, (2003) Displaying method in on-board display – route guidance method based on sex differences. In: Proceedings of IEEE intelligent transportation systems council, Columbus, OH, pp 357–360
8. Ross T, May AJ, Grimsley PJ (2004) Using traffic light information as navigational cues: implications for navigation system design. Transp Res Part F 7:119–134
9. May AJ, Ross T (2006) Presence and quality of navigational landmarks: effect on driver performance and implications for design. Hum Factors 48:346–361
10. Kishi K, sugiura S (1993) Human factors considerations for voice route guidance. In: SAE SP-964 automotive display systems and IVHS. Society of Automotive Engineers, Warrendale, pp 99–109
11. Yoshioka M, Akamatsu M, Matsuoka K, Sato T (2003) Aim of "behavior-based human environment creation technology". In: Proceedings of the XVth triennial congress of the international ergonomics association, Seoul, CD-ROM
12. Vogel K (2002) What characterizes a "free vehicle" in an urban area? Transp Res Part F 5:15–29
13. Liu BS (2007) Association of intersection approach speed with driver characteristics, vehicle type and traffic conditions comparing urban and suburban areas. Accid Anal Prev 39:216–223
14. Akamatsu M, Okuwa M, Onuki M (2001) Development of hi-fidelity driving simulator for measuring driving behavior. J Robot Mechatron 13:409–418
15. Kappe B, Winsum W, Wolffelaar P (2002) A cost-effective driving simulator. In: Proceedings of DSC (driving simulation conference) 2002. INRETS-RENAULT, Paris, pp 123–134
16. The National Advanced Driving Simulator (2010) The University of Iowa. http://www.nads-sc.uiowa.edu/media/pdf/NADS_2010Overview.pdf. Accessed 19 May 2011
17. Akamatsu M, Onuki M (2008) Trends in technologies for representing the real world in driving simulator environment. Rev Automot Eng 29:611–618
18. Pinto M, Cavallo V, Ohlmann T, Espie S, Roge J (2004) The perception of longitudinal accelerations: what factors influence braking maneuvers in driving simulators?. In: Proceedings of DSC (driving simulation conference) 2004 Europe. INRETS-RENAULT, Paris, pp 139–151
19. Kemeny A, Panerai F (2003) Evaluating perception in driving simulation experiments. Trends Cogn Sci 7:31–37
20. Dingus TA, Klauer SG, Neale VL, Petersen A, Lee SE, Sudweeks J, Perez MA, Hankey J, Ramsey D, Gupta S, Bucher C, Doerzaph ZR, Jermeland J, Knipling RR (2006) The 100-car naturalistic driving study, phase 2 – results of the 100-car field experiment, Report no. DOT HS 810 593
21. Akamatsu M, Sakaguchi Y, Okuwa M, Kurahashi T, Takiguchi K (2003) Database of driving behavior measured with equipped vehicles in the real road environment. In: Proceedings of the XVth triennial congress of the international ergonomics association, Seoul, CD-ROM
22. Sato T, Akamatsu M (2006) Analysis of drivers' behaviour before making a right turn at an intersection based on driving behavior data on an actual road. RevAutomot Eng 27:287–294
23. Sato T, Akamatsu M, Daimon T (2006) Investigation on drivers' behaviour while approaching an intersection: comparison between real and simulated driving. In: Proceedings of driving simulation conference Asia/Pacific (DSC-AP 2006), Tsukuba, CD-ROM
24. Sato T, Akamatsu M (2007) Influence of traffic conditions on driver behaviour before making a right turn at an intersection: analysis of driver behaviour based on measured data on an actual road. Transp Res Part F 10:397–413
25. Brackstone M, McDonald M (2007) Driver headway: how close is too close on a motorway? Ergonomics 50:1183–1195
26. Ranney TA (1999) Psychological factors that influence car-following and car-following model development. Transp Res Part F 2:213–219
27. Sato T, Akamatsu M (2009) Analysis of drivers' preparatory behavior before turning at intersections. IET Intell Transp Syst 3:379–389
28. Land MF, Lee DN (1994) Where we look when we steer. Nature 369:742–744
29. Schumacker RE, Lomax RG (2004) SEM Basics. In: Riegert D (ed) A beginner's guide to structural equation modeling. Lawrence Erlbaum Associates, Mahwah, pp 61–78
30. Hu L, Bentler PM (1995) Evaluating model fit. In: Hoyle R (ed) Structural equation modeling: issues, concepts, and applications. Sage, Newbury, pp 76–99
31. Sato T, Akamatsu M (2008) Modeling and prediction of driver preparations for making a right turn based on vehicle velocity and traffic conditions while approaching an intersection. Transp Res Part F 11:242–258
32. Bollen KA (1989) Confirmatory factor analysis. In: Structural equations with latent variables. Wiley, New York, pp 226–318
33. Akamatsu M, Sakaguchi Y, Okuwa M (2003) Modeling of driving behavior when approaching an intersection based on measured behavioral data on an actual road. In: Proceedings of the human factors and ergonomics society 47th annual meeting, Denver, CL, pp 1895–1899
34. Pentland A, Liu A (1999) Modeling and prediction of human behavior. Neural Comput 11:229–242

Books and Reviews

Cacciabue C (2007) Modelling driver behavior in automotive environments: critical issues in driver interactions with intelligent transport systems. Springer, New York

Fuller R, Santos JA (2002) Human factors for highway engineers. Pergamon, Oxford, UK

Shinar D (2007) Traffic safety and human behavior. Emerald, Bingley

Takeda K, Hansen JHL, Erdogan H, Abut H (2009) In-vehicle corpus and signal processing for driver behavior. Springer, New York

Driver Characteristics Based on Driver Behavior

Jianqiang Wang[1], Lei Zhang[1], Xiaojia Lu[2], Keqiang Li[1]
[1]State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing, China
[2]China Agricultural University, Beijing, China

Article Outline

Glossary
Definition of the Subject
Introduction
Driver Behavior Questionnaire Design
Data Statistical Analysis and Quantification Analysis
Comparison and Discussion
Future Directions
Acknowledgments
Bibliography

Glossary

Driver Assistance System (DAS) Also called ADAS (Advanced Driver Assistance Systems), can prevent accidents caused by human errors or alleviate the damage by taking control of the vehicle just before an accident happens. It can contribute to improving traffic safety, environmental impact, energy efficiency, traffic homogeneity, and driver comfort and convenience. An adaptive Driver Assistance System ensures safe driving by assisting the driver in the following ways: reducing driver fatigue and maintaining driver performance by supporting a driver's "recognitions," "judgments," and "actions" displaying warning signs in the case that the actions are judged to be dangerous and taking control of the vehicle in the case that the driver is unable to avoid a collision.

Driver Characteristics The characteristics performed when driving, including the driver's operations to the vehicle (e.g., car-following and lane-changing), the physiological characteristics (e.g., response, cognition, and fatigue), and the psychological characteristics and so on. It has been studied mainly in two main aspects, i.e., driving skill and driving style.

Driver Behavior Questionnaire (DBQ) One of the most widely implemented measurement scales to examine self-reported driving behaviors, developed by Reason et al.

Factor Analysis Factor analysis was invented by psychologist Charles Spearman. It is a collection of methods used to examine how underlying constructs influence the responses on a number of measured variables. It is related to the principal component analysis but not identical; it estimates how much of the variability is due to common factors. It is performed by examining the pattern of correlations (or covariances) between the observed measures. Measures that are highly correlated (either positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

Human Factors It is a theory that studies the physical and mental state of human beings when given different environments, products, and services.

Reliability It is the quality of being dependable or reliable. And it is defined as the proportion of the real variance in the total variance in the statistics. At present, the Cronbach's alpha reliability coefficient method is a commonly used measure. It is specially used for the reliability estimation of the accumulative Likert scale. It is defined as: the ratio of the sum variance of each item and the total variance of the whole scale.

***T*-Test** The *T*-test method is a common method for hypothesis testing. It is used to determine whether there are differences between the two samples (such as A and B) with an unknown but equal population variance. It can calculate the values simply, and is suitable for small sample testing.

Validity It is defined as the ratio of the variance of the scores that related to the measure purposes and the total variance. The content validity and discriminate validity are the two main indexes.

Definition of the Subject

According to the road traffic statistics and analysis information, most accidents can be directly attributed

to the driver mistake and violation behavior. In driving behavior research, the driving skill and style can be seen as the two main components of human factors [1]. The deficient driving skill and aggressive driving style may both lead to serious crash risk. Therefore, different kinds of Advanced Driver Assistance Systems based on intelligent technologies are developed by automotive manufacturers to increase driver skill and decrease driver workload, such as Forward Collision Warning System and Adaptive Cruise Control System [2]. The governments also conduct much work on the driver education and regulation to avoid aggressive behaviors. In order to optimize the effects of technical driving assistance and driver normalization, understanding and measuring of driver skill and style have become important research topics. However, because of the complication and variability of human drivers, it is challenging to convert the descriptive concepts of skill and style to quantitative mathematic variables.

Introduction

A self-reported survey is an effective method for studying the driver behavior, and it can be used to analyze the driver behavior, especially the relationship between abnormal behavior and driver characteristics. A number of driver measurement scales have been designed to investigate the driver characteristics and behavior, such as the Driver Anger Scale [3] and the Driving Skill Inventory [4]. In these methods, the Driver Behavior Questionnaire (DBQ) developed by Reason et al. [5] has recently become one of the most widely implemented measurement scales to examine self-reported driving behaviors. Based on this DBQ and its improved editions, many researches on driver behavior were carried out from different perspectives, e.g., the culture comparison [4], region diversity [6], and the gender and age influences [7]. Although the self-reported survey may affect the data of behavior scale by some subjective factors and objective factors (i.e., the uncertain factors from the outside), it cannot reflect the driver characteristics absolutely [8]. The existing research results show that this method has obvious advantages than previous methods, because a great quantity of behavior data about abnormal driving that cannot be obtained by experiments can be collected quickly and effectively. Meanwhile, the statistical methods that are involved in this paper (e.g., factor analysis, correlation analysis, and cluster analysis) provide an effective tool and basis for driver characteristics analysis. Thus, the indicated factors extracted from the behavior scale can accurately describe the abnormal driver characteristics of driver individual and population to a certain extent.

In this research, the DBQ is applied to explore a quantification method of driver characteristics. A self-reported survey based on DBQ is implemented with 33 Chinese driver participants. Factor analysis is used as the data processing method and a two-dimensional factor structure is established to describe the driver's characteristics of driving skill and driving style respectively. Some comparisons are also executed to validate the reasonability of this method.

Driver Behavior Questionnaire Design

Participants

The self-reported data in this survey are collected from 33 Chinese drivers with different occupations. There are 29 males (87.9%) and 4 females (12.1%). The average age of the participants is 45.54 years (S.D. = 9.86), and the age range is from 30 to 62 years. On average the participants hold their driving licenses for 13.21 years (S.D. = 8.29, range 3–38 years). The average driving mileage is 180.8000 km (S.D. = 18.33), and the driving range is from 20,000 to 600,000 km. Seventeen participants (51.5%) have traffic accidental history, and to them the average accident amount is 3.24 (S.D. = 2.77).

Driver Behavior Questionnaire Method

A self-reported survey is designed based on the Driver Behavior Questionnaire [9] and some modifications are made to adjust the items for Chinese traffic and driver conditions. The survey consists of three parts:

1. Individual information: including gender, age, driving experience (in terms of driving years), and traffic accidental history.
2. A self-evaluation on the driving skills and driving style, which uses an estimation method of Social Comparison [8]. The purpose of self-evaluation is to validate the quantification method. The participants were required to indicate on a six-point

scale (0 = never to 5 = nearly all the time) how often they commit each of the abnormal driving behavior. The skill is rated as five-point scales (0 = very strong and 4 = very weak), and so is the driving style (0 = very prudent and 4 = very aggressive).

3. The main body of the survey, including 20 items describing kinds of driver behaviors, which are extracted and transferred from the original questionnaire, as shown in Table 1. According to the definition and analysis of the original DBQ, the first to the tenth items are aggressive violation behaviors. The 11th to the 14th items are fatal errors which may cause serious accidents, and the 15th to the 20th items are just lapses, which are not very dangerous but may bring some troubles.

Driver Characteristics Based on Driver Behavior.

Table 1 Driver behavior questionnaire and statistical analysis results

No	Item	Mean	S.D.	*r* with total score
1	Drive so close to the car in front that it would be difficult to stop in an emergency	0.8485	0.7953	0.471**
2	Cross-junction knowing traffic lights have already turned	0.5152	0.6185	0.289
3	Disregard the speed limit on a residential road	0.9697	0.8833	0.473**
4	Disregard the speed limit on a intercity highway	0.9394	0.8993	0.411*
5	Become angered by another driver and show anger to him	1.6364	1.0252	0.485**
6	Overtake a slow driver on the inside	1.5152	0.7953	0.576**
7	Race away from traffic lights to beat the car beside you	0.5152	0.6185	0.460**
8	Become angered by another driver and give chase	0.8485	0.7550	0.524**
9	Sound your horn to indicate your annoyance to another road user	1.3636	0.7424	0.414*
10	Stay in a closing lane and force your way into another	0.5758	0.6629	0.439*
11	Fail to notice pedestrians are crossing in your path of traffic	0.5455	0.6170	0.367*
12	Fail to check your rearview mirror before pulling out, changing lanes, etc.	0.5758	0.7513	0.438*
13	Underestimate the speed of an oncoming vehicle when overtaking	1.0303	0.8095	0.321
14	Brake too quickly on a slippery road or steer the wrong way in a skid	0.5152	0.5658	0.473**
15	Get into the wrong lane approaching a roundabout or a junction	1.2121	0.7398	0.493**
16	Misread the signs and exit from a roundabout on the wrong road	1.2424	0.9024	0.692**
17	Forget where you left your car in a car park	1.2121	1.0828	0.587**
18	Hit something when reversing that you had not previously seen	1.0303	0.6840	0.460**
19	Intending to drive to destination A, find yourself on the road to destination B	0.5758	0.6629	0.567**
20	Have no clear recollection of the road along which you have just been traveling	1.0303	0.9180	0.444**
Total score		18.6970	7.3802	1.000

*$p < 0.05$

**$p < 0.01$

Questionnaire Validity and Reliability Analysis

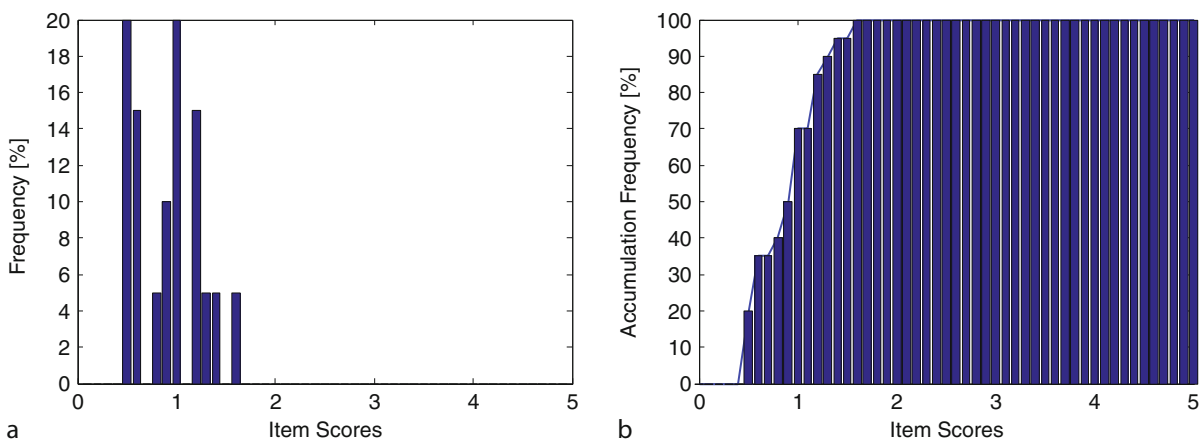
In the statistical field, reliability, and validity are the two main indexes which can determine whether the results of such self-reported questionnaire are reliable and accurate. A questionnaire with high reliability and validity can be considered as the standard scale to test and quantify some characteristics of the participants [10]. In this section, reliability and validity have been analyzed and verified based on the simplified DBQ.

In statistics, reliability is defined as: the proportion of the real variance in the total variance [10]. When the scale has reliability larger than 0.6, it is persuasive. In general, the reliability of scale is estimated based on the sample values that measured, because the variance is hard to accurately calculate. At present, the Cronbach's alpha reliability coefficient method is the most commonly used method [11]. This method is specially used for the reliability estimation of the accumulative Likert scale. It is defined as: the ratio of the sum variance of each item and the total variance of the whole scale. According to the items shown in Table 1 and the total variance, the Cronbach's alpha reliability coefficient of the simplified questionnaire can be obtained as 0.8545, which indicates that the reliability of this scale is very high. Therefore, this scale is relatively persuasive.

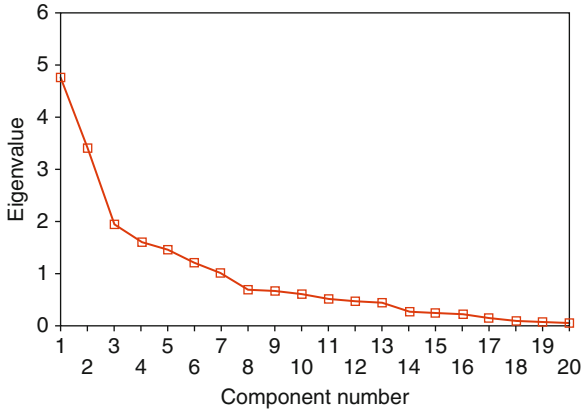
Driver Characteristics Based on Driver Behavior. Table 2 DBQ score analysis of the driver groups

	N	Mean	S.D.	*T* value	*T* Sig.	Mean difference
Male	29	18.310	7.517	−0.806	0.426	−3.190
Female	4	21.500	6.455			
Age $<$ 45 years	18	19.333	7.348	0.537	0.595	1.400
Age $\geq$ 45 years	15	17.933	7.601			
Driving years $<$ 10 years	15	19.733	5.587	0.731	0.47	1.900
Driving years $\geq$ 10 years	18	17.833	8.665			
No accidental history	16	15.750	7.344	2.382	0.024*	−5.721
With accidental history	17	21.471	6.443			

*$p < 0.05$



Driver Characteristics Based on Driver Behavior. Figure 1
Statistical results of item score frequency. (**a**) Frequency distribution and (**b**) accumulation frequency distribution



Driver Characteristics Based on Driver Behavior.
Figure 2
Scree plot of factor analysis

Driver Characteristics Based on Driver Behavior.
Table 3 Factor loading matrix

Item No.	Behavior	Factor 1	Factor 2
16	Lapse	0.795	
20	Lapse	0.749	
19	Lapse	0.746	
11	Error	0.641	
14	Error	0.638	
18	Lapse	0.574	
17	Lapse	0.554	
15	Lapse	0.507	
13	Error	0.469	
12	Error	0.466	
7	Violation		0.804
8	Violation		0.693
4	Violation		0.682
3	Violation		0.640
9	Violation		0.612
6	Violation		0.609
1	Violation		0.595
2	Violation		0.516
5	Violation		0.366
10	Violation		0.350

Validity is mainly used to express the ability that the scale is able to truly measure the desired information, i.e., the effective degree of the scale. It is defined as the ratio of the variance of the scores that related to the measure purposes and the total variance. In this section, the content validity and discriminate validity of the DBQ (the two items of the validity) are assessed by mainly utilizing the correlation analysis and average analysis method.

The content validity analysis can be used to demonstrate the validity of the scale by the correlation analysis of each item score and the total score. The more the correlation between each item score and the total score, the more the items involved in the scale reflect the same theme. The measure index of content validity is defined as the Pearson correlation coefficient (in terms of r_{xy}) between each item score and the total score, whose absolute value is less than 1. The judging rule is that: the closer $|r_{xy}|$ approaches to 1, the more significant the linear relation is. By analyzing the content validity of the DBQ in this paper, it can be seen that the correlation coefficients are all positive. In addition, 90% of the items have high positive interrelations with the total score under 0.05 and 0.01 significance levels. Therefore, this DBQ has a good content validity.

Discriminate validity is expressed as the ability of discriminating the characteristics of the measured sample from the scale. In this paper, it is desired that the scale can be used to analyze the driver characteristics, especially the driver behavior characteristics in traffic accidents. Therefore, the participants are divided into groups based on gender, age, driving years, mileage, and accidental history respectively, and in order to analyze the discriminate validity of the DBQ, *T*-test is carried out to investigate the difference between the groups. As Table 2 shown, the significant probability between groups divided by the accidental history is less than 0.05, which indicates that there is significant difference. Meanwhile, the total score of the group with accidental history is obviously higher than the group with no accidental history. It indicates that the participants having higher frequency behaviors are more likely to be involved in accidents. In addition, in groups divided by gender, age, and driving years, there are some differences between the average scores: the average score of the low age group is higher than that of the high age group, and average score of the low driving

Driver Characteristics Based on Driver Behavior.

Table 4 The factor score coefficient

Item No.	*a*	*b*
1	−0.001	0.154
2	−0.040	0.140
3	−0.014	0.168
4	−0.054	0.185
5	0.050	0.087
6	0.034	0.153
7	−0.061	0.218
8	0.002	0.180
9	−0.018	0.161
10	0.060	0.082
11	0.163	−0.063
12	0.112	0.001
13	0.119	−0.043
14	0.155	−0.010
15	0.117	0.033
16	0.188	0.022
17	0.125	0.059
18	0.139	−0.010
19	0.181	−0.013
20	0.189	−0.066

years is higher than that of the high driving years' group, but the difference is not significant. The average analysis indicates that the DBQ have a significant discriminate validity in the aspect of accidental history.

Data Statistical Analysis and Quantification Analysis

Data Statistical Analysis

The item scores of the questionnaire were analyzed, and the mean score, the standard deviation of the tested samples and that of the total scores are shown in Table 1. Figure 1 is the statistical results of item score frequency. It can be seen that the scores of the items are generally low, and almost all less than 2, more than 50% of these less than 1, which indicates that most of the abnormal driving behaviors in this DBQ are rare. The main reason for this phenomenon may be that most of the participants have high education background and good driving experience.

Quantification Analysis

Factor Analysis Factor analysis theory is applied to the data processing and extraction of this 20-item DBQ. Principle components analysis with oblique rotation is implemented as the factor analysis method to investigate the factor structure of the DBQ. A two-dimensional factor structure is established, too. This structure is also suggested in the scree plot shown in Fig. 2, and the curve becomes smooth from the third component. The two factors could account for 40.1% of the total variance. Table 3 shows the factor load matrix and the items with load less than 0.35 are omitted.

This factor analysis method distinguishes the different types of behaviors clearly. The first factor accounts for 20.8% of the total variance and contains ten items which are all lapse and error behaviors. The second factor contains the other ten items which describe violation behaviors. It could be assumed that both of the error and lapse are attributed to the driver's unintentional passive characteristics caused by the lack of driving skill. Taking the 12th item "Fail to check your rearview mirror before pulling out, changing lanes, etc.," as an example: the behavior occurs because the driver is short of training on this "mirror-checking" experience. As viewed from the behavior's property, this behavior is similar to the lapse behavior such as the 18th item "Hit something when reversing that you had not previously seen," and the diversities between these items are only the different driving skills presented by the behaviors. Therefore, the first factor could be defined as F_{skill} to describe the driver's characteristics on driving skills. The violation behaviors indicate the driver's intentional active characteristics attributed to driving style. The more aggressive style the driver is inclined to, the more violation behaviors may occur during his/her driving. Similarly, the second factor is defined as F_{style} to describe the driver's style characteristics. Based on the principle of factor analysis, these two factors are independent and represent the human driver's unintentional (not purposive) and intentional (purposive) behaviors respectively.

Driver Characteristics Quantification The factor analysis converts the 20 items to two factors F_{skill} and F_{style}, and provides a method to quantify individual driver's characteristics of driving skill and style based on the scale point data of self-reported questionnaire. The quantification of F_{skill} and F_{style} is calculated as the equation below. X_1, X_2,⋯,X_M are the standardized scale points of the DBQ items provided by the participants during self-reported survey, and a_1, $a_2 \cdots a_M$, b_1, $b_2 \cdots b_M$ are the factor score coefficients (M =20).

$$F_{skill} = a_1X_1 + a_2X_2 + \cdots + a_MX_M$$
$$F_{style} = b_1X_1 + b_2X_2 + \cdots + b_MX_M$$

Based on the original data of the 33 participants, the coefficients a_1, $a_2 \cdots a_M$, b_1, $b_2 \cdots b_M$ could be estimated with regression analysis and the results are shown in Table 4.

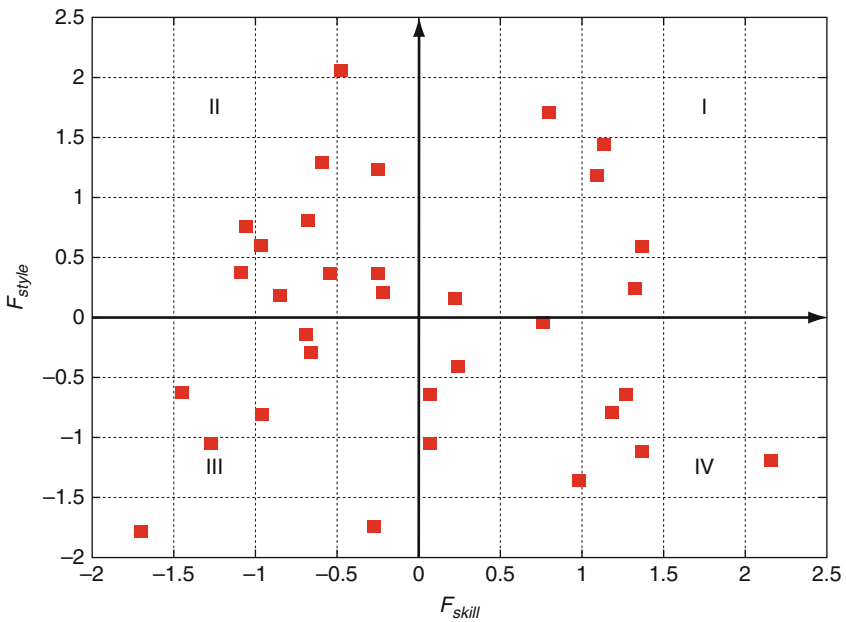
The quantified factors are normalized that the mean is 0 and the standard deviation is 1. The drivers with larger driving skill factor F_{skill} have higher frequency of error and lapse behaviors, and the larger value of driving style factor F_{style} means more violation behaviors. The data points of the 33 participants' characteristics factors distribute in all of the four quadrants, which are shown in Fig. 3. There are six participants (18.2%) in the first quadrant, 11 participants (33.3%) in the second quadrant, seven participants (21.2%) in the third quadrant, and nine participants (27.3%) in the fourth quadrant. This proportion of distribution indicates that the drivers are inclined to two main tendencies: one is aggressive style with better skill (quadrant II), and the other is prudent style with worse skill (quadrant IV). This result is coincident with the human psychological features.

Comparison and Discussion

The participants' self-evaluations on driving skill and style are collected during the survey, and the statistical results are shown in Table 5. According to the five-point scales, most participants are confident with

Driver Characteristics Based on Driver Behavior.
Table 5 Driver self-evaluation statistical results

	Skill	Style
Mean	1.212	1.879
S.D.	0.600	1.023



Driver Characteristics Based on Driver Behavior. Figure 3
Factor score distribution

Driver Characteristics Based on Driver Behavior. Table 6 Correlative coefficients analysis results of factor score and self-evaluation

	Self-evaluation of driving skill	Self-evaluation of driving style	Driving skill factor F_{skill}	Driving style factor F_{style}
Self-evaluation of driving skill	1	$r = -0.262$	$r = -0.542^{**}$	$r = -0.294$
		Sig = 0.140	Sig = 0.001	Sig = 0.096
Self-evaluation of driving style	–	1	r = 0.225	$r = 0.481^{**}$
			Sig = 0.208	Sig = 0.005
Driving skill factor F_{skill}	–	–	1	$r = 0.000$
				Sig = 1
Driving style factor F_{style}	–	–	–	1

$^{**}p < 0.01$

Driver Characteristics Based on Driver Behavior. Table 7 F_{skill} analysis of the driver groups

	N	Mean	S.D.	*T* value	*T* Sig.	Mean difference
Male	29	−0.129	0.987	−2.101	0.044^{**}	−1.065
Female	4	0.936	0.487			
Age < 45 years	18	0.047	0.924	0.290	0.774	0.103
Age ≥ 45 years	15	−0.056	1.115			
Driving years < years	15	0.231	0.894	1.222	0.231	0.424
Driving years ≥ 10 years	18	−0.192	1.067			
No accidental history	16	−0.301	0.971	−1.726	0.094^{*}	−0.584
With accidental history	17	0.283	0.970			

$^{*}p < 0.1$
$^{**}p < 0.05$

Driver Characteristics Based on Driver Behavior. Table 8 F_{style} analysis of the driver groups

	N	Mean	S.D.	*T* value	*T* Sig.	Mean difference
Male	29	0.049	0.975	0.754	0.456	0.405
Female	4	−0.356	1.263			
Age < 45 years	18	0.102	0.950	0.633	0.532	0.223
Age ≥ 45 years	15	−0.122	1.077			
Driving years < 10 years	15	−0.034	0.832	−0.177	0.860	−0.063
Driving years ≥ 10 years	18	0.029	1.144			
No accidental history	16	−0.229	0.893	−1.287	0.208	−0.444
With accidental history	17	0.215	1.073			

their driving skill and consider themselves as prudent drivers.

For the purpose of investigating the relationships between the driver background information, self-evaluation and the quantified factor scores, correlation analysis of these data is carried out. Table 6 shows the correlative coefficients analysis results of factor score and self-evaluation.

It can be seen that there are all positive correlations that with the significance level less than 0.01 between the driving skill factor F_{skill} and driving skill self-evaluation, the driving style factor F_{style} and driving style self-evaluation. These results show that the quantified factor scores are in accordance with the driver's self-evaluation.

The comparisons between the groups divided by gender, age, driving years, and accidental history are also conducted with *T*-test of F_{skill} and F_{style}, which are shown in Tables 7 and 8 respectively.

It can be seen that there are certain differences between the average values of common factors with the different groups. As for the group divided by gender, the average values of the driving skill factor F_{skill} of the female group are all higher than those of the male group. And the difference significant level is lower than 0.05, which indicates that there are certain differences for the driving ability of female participants compared with that of male participants. At the same time, the average values of driving style factor F_{style} with the female group are all lower than those of the male group, which indicates that the driving style of the female participant tends to be more conservative compared with the male participant. It accords with the gender characteristics to a certain extent.

As for the group divided by age and driving years, the average values of the driving skill factor F_{skill} of the low age group and the short driving years group is higher than those of the high age group and the long driving years' group. In addition, the average values of the driving style factor F_{style} with the low age group is higher than those of the high age group, and there is almost no difference between the short driving years group and the long driving years group. Then the difference also accords with the conventional understanding, but there is no significant difference among the most average values of the driving skill factor F_{skill}, and the same as the driving style factor F_{style}.

As for the group divided by accidental history, the average values of the driving skill factor F_{skill} with the no accidental history group are lower than those of the accidental history group and the significant level is lower than 0.1, which is an obvious difference. The above analysis indicates that the driver characteristics and the difference of the accident history can be reflected by the common factor. And the driver with the high driving skill factor F_{skill} or driving style factor F_{style} is much easier to have an accident.

Based on the comparisons between quantified factor scores and driver self-evaluation, the group analysis, the reasonability of the quantification method is validated that the quantified factors could embody the driver's characteristics tendencies on skill and style.

Future Directions

In this paper, a quantification method is explored to investigate the human driver behavior characteristics. Thirty-three Chinese drivers participate in a self-reported survey based on Driver Behavior Questionnaire (DBQ) and a two-dimensional factor structure is established from the original data with factor analysis theory.

The extracted F_{skill} and F_{style} factors can be calculated by the data of driver self-reported survey expediently, which describe the driver's unintentional driving skill and intentional driving style characteristics, respectively. Some comparisons and discussions are also carried out to validate this method. The quantification method of driving skill and driving style could be applied to the development of adaptive Driver Assistance Systems, e.g., providing more assistance to drivers with less driving skill and effective warning to the drivers with aggressive driving style. The quantified evaluation of drivers could also be utilized to train and educate drivers and help to improve the driving skill and correct the driving style. Meanwhile, the comparative analysis of the DBQ and real road experimental data can be carried out in the future work.

Acknowledgments

The project was supported by NSFC (No. 50705047). The authors especially thank Mr. Qing Xiao (China Agricultural University) and Mrs. Furui Yang (China Agricultural University) for their contribution to the preparation of the chapter.

Bibliography

1. Elander J, West R, French D (1993) Behavioral correlates of individual differences in road traffic crash risk: an examination of methods and findings. Psychol Bull 113:279–294
2. Aufrere R, Gowdy J, Mertz C, Thorpe C, Wang C, Yata T (2003) Perception for collision avoidance and autonomous driving. Mechatronics 13(10):1149–1161
3. Deffenbacher JL, Oetting ER, Lynch RS (1994) Development of a driving anger scale. Psychol Rep 74:83–91
4. Lajunen T, Parker D, Summala H (2003) The Manchester driver behaviour questionnaire: a cross-cultural study. Accid Anal Prev 942:1–8
5. Reason JT, Manstead ASR, Stradling SG, Baxter JS, Campbell K (1990) Errors and violations on the road: a real distinction? Ergonomics 33:1315–1332
6. Xie C, Parker D (2002) A social psychological approach to driving violations in two Chinese cities. Transport Res Part F: Traffic Psychol Behav 5(4):293–308
7. Ozkana T, Lajunen T, Summala H (2006) Driver behaviour questionnaire: a follow-up study. Accid Anal Prev 38:386–395
8. Lajunen T, Summala H (2003) Can we trust self-reports of driving? Effects of impression management on driver behavior questionnaire responses. Transp Res F 6:97–107
9. Davey J, Wishart D, Freeman J, Watson B (2007) An application of the driver behaviour questionnaire in an Australian organisational fleet setting. Transp Res F 10:11–21
10. Huixin K, Shen H (2005) The statistical analysis of the research, 2nd edn. The Press of China Media University, Beijing
11. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16(3):297–334

Driver Inattention Monitoring System for Intelligent Vehicles

Yanchao Dong[1], Zhencheng Hu[2]
[1]Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan
[2]Graduate School of Science and Technology, Computer Science Department, Kumamoto University, Kumamoto, Japan

Article Outline

Glossary
Definition of the Subject
Introduction
Distraction and Fatigue Effects on Driving Behavioral Performance
Commercial Products and Activities for Driver Inattention Detection
Current Methods to Detect Driver Inattention
Discussion: Future Directions
Conclusion
Bibliography

Glossary

DIMS Driver inattention monitoring system, to monitor the attention status of the driver.

Distraction Driver distraction is a diversion of attention away from activities critical for safe driving toward a competing activity.

Driver biological Utilize driver biological signals, e.g., electroencephalography (EEG), electrocardiogram (ECG), electro-oculography (EOG), surface electromyogram (sEMG), to estimate driver attention status.

Driver physical Utilize driver's physical signals, e.g., eye closure duration, blink frequency, nodding frequency, fixed gaze, and frontal face pose, to estimate driver attention status.

Driving performance Utilize driving performance, e.g., pressure distribution on the seat, car-following, steering wheel angle, accelerator pedal position, lane boundaries, and upcoming road curvature, to estimate driver attention status.

Fatigue Driver fatigue refers to a combination of symptoms such as impaired performance and a subjective feeling of drowsiness.

Hybrid Combining driver physical measures with driving performance measures to estimate driver attention status.

Inattention Driver inattention represents diminished attention to activities that are critical for safe driving in the absence of a competing activity.

Measures The way to estimate the driver's attention status.

Physical signal extraction The approaches for extracting driver physical signals.

Subjective report Subjective self-assessment of attention status.

Definition of the Subject

Driver inattention is a major factor in highway crashes. The National Highway Traffic Safety Administration (NHTSA) estimates that approximately 25% of police-reported crashes involve some forms of driver inattention – the driver is distracted, asleep or fatigued, or otherwise "lost in thought" [1]. This entry reviews

the state-of-the-art technologies for monitoring driver inattention, which can be classified into two main categories: distraction and fatigue. Driver inattention is a major factor in most traffic accidents. Research and development has been actively carried out for decades with the goal of precisely determining the drivers' state of mind. This entry summarizes these approaches by dividing them into five different types of measures:

1. Subjective report measures
2. Driver biological measures
3. Driver physical measures
4. Driving performance measures
5. Hybrid measures

Among these approaches, subjective report measures and driver biological measures are not suitable under real driving conditions, but could serve as some rough ground truth indicators. The hybrid measures are believed to give more reliable solutions compared with single driver physical measures or driving performance measures, because the hybrid measures minimize the number of false alarms and maintain a high recognition rate, which promote acceptance of the system. Also, a discussion on some nonlinear modeling techniques commonly used in the literature is made.

Introduction

Driver inattention is a major factor in highway crashes. The National Highway Traffic Safety Administration (NHTSA) estimates that approximately 25% of police-reported crashes involve some forms of driver inattention – the driver is distracted, asleep or fatigued, or otherwise "lost in thought" [1]. A common definition of driver inattention is given in [2]: "Driver inattention represents diminished attention to activities that are critical for safe driving in the absence of a competing activity."

A study by the AAA FTS (American Automobile Association Foundation for Traffic Safety) utilized five categories for the driver attention status: attentive, distracted, looked but did not see, sleepy, and unknown [3]. The category of looked but did not see can be considered to be a kind of cognitive distraction, and the word "sleepy" could be replaced by the more comprehensive word "fatigued." In this entry, two categories of inattention are proposed: *distraction* and *fatigue*.

The causes of driver distraction are diverse and pose large risk factors – over half of the crashes involving inattention were caused by driver distraction [1, 2]. After an intensive study on the various definitions of driver distraction appeared in the literature, a more general definition is proposed by Lee et al. [2]: "Driver distraction is a diversion of attention away from activities critical for safe driving toward a competing activity."

Thirteen types of potentially distracting activities are listed in [3]: eating or drinking, outside person, object or event, talking or listening on cellular phone, dialing cellular phone, using in-vehicle-technologies, etc. Because the distracting activities take many forms, the NHTSA classifies distractions into four categories from the viewpoint of the driver's functionality: visual distraction (e.g., looking away from the roadway), cognitive distraction (e.g., being lost in thought), auditory distraction (e.g., responding to a ringing cell phone), and biomechanical distraction (e.g., manually adjusting the radio volume) [1]. Two more categories are added by Lee et al. [2]: olfactory distraction and gustatory distraction. Many distracting activities can involve more than one of these components (e.g., talking to a phone while driving creates a biomechanical, auditory, and cognitive distraction).

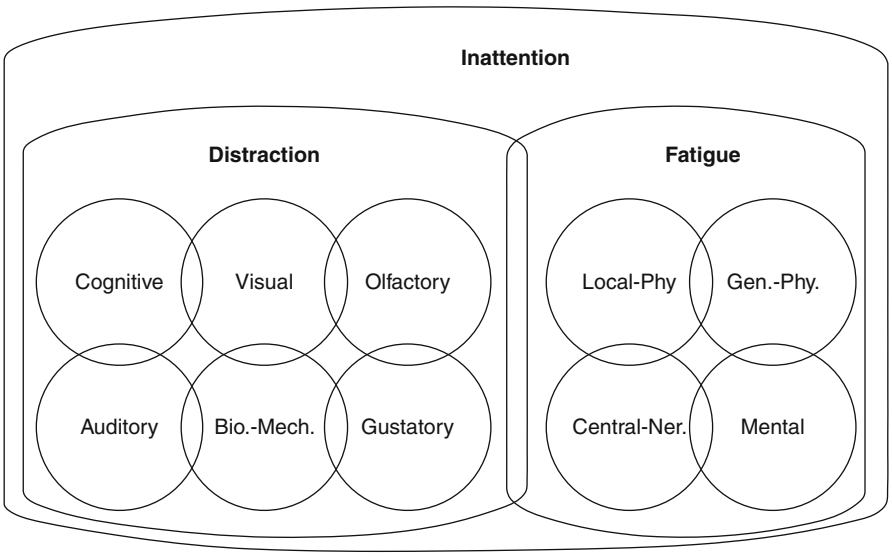
The phenomenon of fatigue is different from that of distraction. The term "fatigue" refers to a combination of symptoms such as impaired performance and a subjective feeling of drowsiness [4]. Even with the intensive research that has been performed, the term "fatigue" still does not have a universally accepted definition [5]. Thus, it is difficult to determine the level of fatigue-related accidents. However, studies show that 25–30% of driving accidents are fatigue related [6]. In their definition, the European Transport Safety Council (ETSC) states that fatigue "concerns the inability or disinclination to continue an activity, generally because the activity has been going on for too long" [7]. From the viewpoint of individual organ functionality, there are different kinds of fatigue such as local physical fatigue (e.g., in a skeletal or ocular muscle), general physical fatigue (following heavy manual labor), central nervous fatigue (sleepiness), and mental fatigue (not having the energy to do anything). Central nervous fatigue and mental fatigue are the most dangerous types for driving, because these will

eventually lead to sleepiness, increasing the probability of an accident.

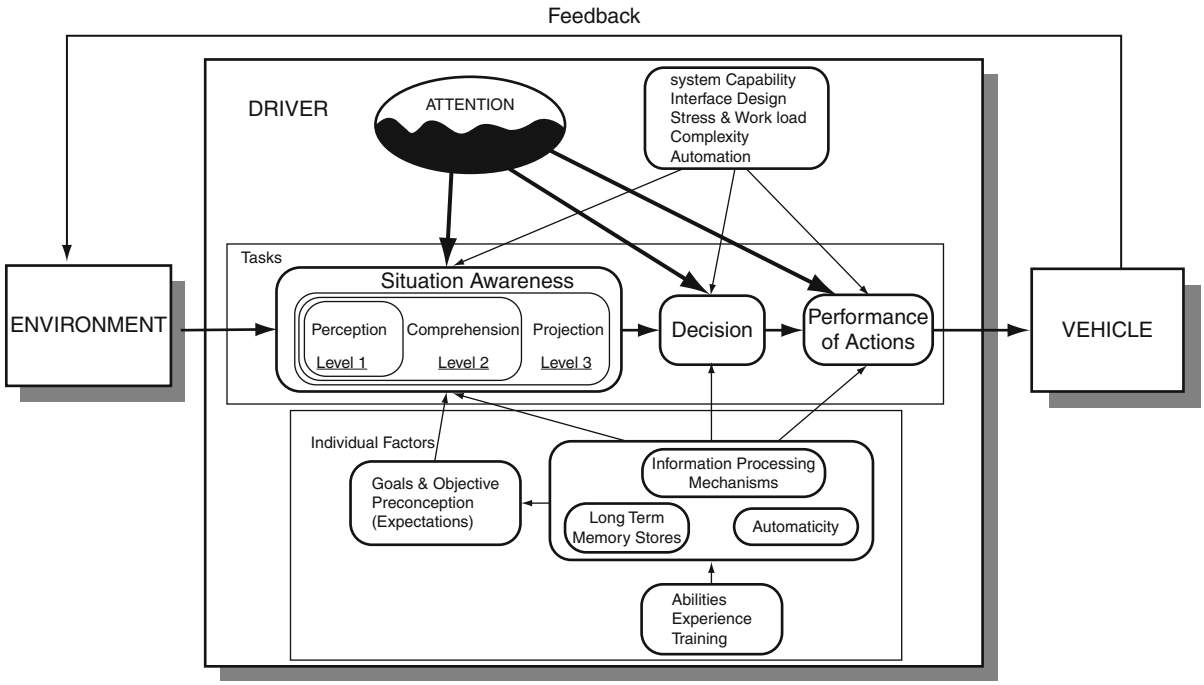
The ETSC defines four levels of sleepiness based on behavioral terms [7]: (a) completely awake, (b) moderate sleepiness, (c) severe sleepiness, and (d) sleep. In an attempt to avoid having an accident, most sleepy drivers will try to fight against sleep with different durations and sequences of the physiological events that precede the onset of sleep [8]. When a driver becomes fatigued and begins to fall asleep, the following symptoms can be observed: repeated yawning, confusion and thinking seems foggy, feeling depressed and irritable, slower reaction and responses, daydreaming, difficulty keeping eyes open and burning sensation in the eyes, lazy steering, difficulty maintaining concentration, swaying of head or body from nodding off, vehicle wanders from the road or into another lane, nodding off at the wheel, breathing becomes shallow, heart races, etc. Different individuals show different symptoms and the degrees vary. Thus, there is no concrete method to measure the level of fatigue. The ETSC's study [7] showed that the level of fatigue or sleepiness (sleepiness is the outside exhibition of fatigue) is a function of the amount of activity in relation to the brain's physiological waking capacity. Several factors can influence this physiological waking capacity, and hence lower the fatigue threshold [4, 5, 7, 9] such as disturbed sleep, the low point in the circadian rhythm, hard work prior to driving, etc. These factors are independent of the activity being undertaken, but result in the fatigue effect of that activity appearing more quickly. Thus, fatigue cannot be seen simply as a function of the duration of time engaged in driving.

Driving inattention includes two main categories and each of them includes a few subcategories, as shown in Fig. 1. The purpose of driving inattention monitoring system (**DIMS**) is to monitor the attention status of the driver and if driving inattention is detected some measures should be done to make the driving safe, depending on the inattention type and level.

Driving is a process involving situation awareness of the environment, decision making, and the performance of actions, as shown in Fig. 2 [10]. In this process, the most complicated stage is the situation awareness. In [10], a three-level situation awareness model is defined as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." The deployment of attention in the perception process acts to present certain constraints on a person's ability to accurately perceive multiple items in parallel, and is a major limitation on situation awareness. Direct attention is needed, not only for perceiving and processing the



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 1
Inattention categories



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 2
Information processing and attention [10]

available cues, but also in the later stages of decision making and response execution. In a complex and dynamic driving environment, attention demands result from information overload, complex decision making, and the performance of multiple tasks. Thus, monitoring the attention status is vital for maintaining safe driving.

The purpose of the driver inattention monitoring system (**DIMS**) is to monitor the attention status of the driver. If driver inattention is detected, different countermeasures should be taken to maintain driving safety, depending on the types and levels of inattention. **DIMS** has been an active research field for decades. The first international conference on driver distraction and inattention was held in 2009 [11]. Some auto companies have already installed some simple function driver fatigue monitoring systems in their high-end vehicles. Yet, there is still a great need to develop a more reliable and fully functional **DIMS** using cost-efficient methods for a real driving context. It is believed that the development of signal processing and computer vision techniques will attract more attention to the study of this field in the coming years. With the intention of benefiting those interested in, or about to enter, this field, this entry gives a comprehensive review of the state of the knowledge on driver inattention. It thus provides a clear view of the previous achievements and the issues that still need to be considered.

The arrangement of this entry is as follows. We introduced the driver inattention concept in section "Introduction". Next, the effects of driver distraction and fatigue on driving performance are presented in section "Distraction and Fatigue Effects on Driving Behavioral Performance". Because some commercial products relative to inattention detection have emerged on the market in recent years, section "Commercial Products and Activities for Driver Inattention Detection" is devoted to reviewing these products. section "Current Methods to Detect Driver Inattention" presents a detailed review of the scientific researches on inattention detection. Five types of measures for inattention detection are presented in this section: (a) subjective report measures, (b) driver biological measures, (c) driver physical measures, (d) driving performance measures, and (e) hybrid measures. After a discussion of future directions in section

"Discussion: Future Directions", a conclusion is given in section "Conclusion".

Distraction and Fatigue Effects on Driving Behavioral Performance

This section concentrates on how distraction and fatigue affect a driver's behavior and driving performance. Exploring these effects could provide useful information for the development of real-time distraction and fatigue detection algorithms.

Effects of Distraction

Performing a cognitively demanding task while driving would influence both the driver's visual behavior and driving performance (as indicated by braking behavior).

Driver Behavior Patterns With an increase in the cognitive demand, many drivers changed their inspection patterns on the forward view. Angell et al. [12] indicated that the eye-glance pattern could be used to discriminate driving while performing a secondary task from driving alone, and could be used to discriminate high- from low-workload secondary tasks. More facts associated with cognitive distraction driving can be found in [13, 14]: Drivers narrowed their inspection of the outward view and spent more time looking directly ahead. They reduced their inspection of the instruments and mirrors, and reduced their glances at traffic signals and the area around an intersection. Rantanen and Goldberg [14] found that the visual field shrank by 7.8% during a moderate-workload counting task and by 13.6% during a cognitively demanding counting task. Drivers had fewer saccades per unit time, which was consistent with a reduction in glance frequency and less exploration of the driving environment, and in some cases drivers shed these tasks completely and did not inspect these areas at all [15]. Hayhoe [16] showed links between eye movement (fixation, saccade, and smooth pursuit), cognitive workload, and distraction. Fixations occur when an observer's eyes are nearly stationary. Saccades are very fast movements that occur when visual attention shifts from one location to another. Smooth pursuits occur when an observer tracks a moving object such as a passing vehicle. Saccade distance decreases as task complexity increases, which indicates that saccades may be a valuable index of mental workload [17]. In contrast, the amount of head movement increased when cognitive loads were imposed. It is believed that this is a compensatory action by which a driver attempts to obtain a wider field of view [18]. Miyaji et al. [18] proposed that the standard deviations of eye movement and head movement could be suitable for detecting the states of cognitive distraction in subjects. Both cognitive and visual distractions caused gaze concentration and slow saccades when drivers looked at the roadway, and cognitive distraction increased blink frequency [19]. Liang and Lee [19] found that visual distraction resulted in frequent, long off-road glances. A report from the Safety Vehicle Using Adaptive Interface Technology (SAVE-IT) program showed that eyes-off-road glance duration, head-off-road glance time, and standard deviation of lane position (SDLP) are good measures of visual distraction [20].

Other Physiological Responses When cognitive loads (conversation or arithmetic) were imposed on subjects, pupil dilation occurred by the acceleration of the sympathetic nerve [18]. The average heart rate also increased by approximately 8 beats per minute. However, the average value of the heart rate RRI decreased under the same situation [18]. Itoh [21] pointed out that performing a cognitively distracting secondary task (e.g., talking, thinking about something, etc.) during driving caused a decrease in the driver's temperature at the tip of the nose, and this effect was reproducible. It was reported in [22] that a considerable and consistent skin temperature increase in the supraorbital region could be observed during cognitive and visual distractions. Berka [23] found that the electroencephalography (EEG) signal also contained information about the task engagement level and mental workload.

Driving Performance Significant changes were observed in a driver's vehicle control as a consequence of performing the additional cognitive tasks while driving. Ranney [24] found that distraction may be associated with lapses in vehicle control, resulting in unintended speed changes or allowing the vehicle to drift outside the lane boundaries. Zhou et al. [25]

found the influences on the lane-changing behavior when a secondary task was being performed, which included a reduction in the frequency of the checking behavior (check a side mirror or speedometer), a delay in the checking behavior, and a longer time for the checking behavior. Carsten and Brookhuis [26] found that the effects of cognitive distraction on driving performance differed considerably from those of visual distraction. Visual distraction affects a driver's steering ability and lateral vehicle control, while cognitive distraction affects longitudinal vehicle control, particularly car-following. Liang and Lee [19] also found that cognitive distraction made steering less smooth, but improved lane maintenance. Liang and Lee [19] found that steering neglect and overcompensation are associated with visual distraction, while under-compensation is associated with cognitive distraction. Overall, visual distraction interferes with driving performance more than cognitive distraction. An apparently anomalous finding is that when secondary task cognitive demands increased, a driver's lateral control ability was found to improve [26]. Harbluk et al. [13, 15] found an increased incidence of hard braking associated with cognitive distraction driving.

Effects of Fatigue

When a driver is fatigued, certain physical and physiological phenomena can be observed. These include changes in brain waves or EEG, eye activity, facial expressions, head nodding, body sagging posture, heart rate, pulse, skin electric potential, gripping force on the steering wheel, and other changes in body activities.

Driver Behavior Patterns Eskandarian et al. [27] found that the follow actions were correlated with fatigue: Drivers exhibited a reflexive head nod after checking the side mirrors; the head motions were significantly less frequent; the number of times drivers touched or scratched their chin, face, head, ears, eyes, and legs significantly increased; drivers were inclined to turn their head to the left to relieve muscular tension in the neck; eye blinking activity radically increased; episodes of yawning were more frequent; and they tended to adopt more relaxed hand positions on the steering wheel. Particularly for eye blinking patterns, PERCLOS [28], the percentage of time the eye is more than 80% closed, is one of the most widely accepted measures in the scientific literature for drowsiness detection. It has been validated using both EEG data and subjective evaluation.

Other Physiological Responses The activity of a low frequency EEG ranging from 0 to 20 Hertz has a significant relationship with sleepiness. The spectral analysis of an EEG shows the transition from wakefulness to sleep can be described as a shift toward slower EEG frequencies. In the alert condition, the appearance of β activity is common in the EEG. α activity is also normally found in the occipital regions (O1 and O2) in the awake and relaxed condition. When a driver gets drowsy, a burst of α activity can often be seen in the central regions of the brain (C3 and C4). However, some people do not show any α activity. As the driver gets drowsier, the α activity is replaced by θ activity. When δ activity occurs in the EEG, the driver is no longer awake; this is an indicator of deep sleep [29].

Driving Performance It has been reported that sleep-deprived drivers have a lower frequency of steering reversals (every time the steering angle crosses zero degrees) [30], a deterioration of steering performance [31], a decrease in the steering wheel reversing rate [32], more frequent steering maneuvers during wakeful periods, no steering correction for a prolonged period of time followed by jerky motion during drowsy periods [33], low velocity steering [34], large amplitude steering wheel movements, and large standard deviations in the steering wheel angle [35]. Zhong et al. [36] found that when drivers had a fatigued status, the steering wheel angle and vehicle tracking became irregular, and the range of deviation greatly increased. Several researchers found that the lane-tracking ability decreased as the time on the task increased [31]. Variables such as the times of lane departures, SDLP, and maximum lane deviation were found to be highly correlated with eye closures [37]. The mean square of lane deviation, mean square of high pass lateral position, and SDLP showed good potential as drowsiness indicators [38].

Dingus et al. [34] found that the yaw deviation variance and the mean yaw deviation (calculated over a 3 min period) showed some promise as drowsiness indicators. However, no strong correlations between

drowsiness and braking or acceleration were found in [34, 39]. Generally, vehicle speed variability has not shown any strong correlation with drowsiness [39]. However, some reports found that the standard deviation of speed increased from the third driving hour with a time interval of 45 min [40].

Commercial Products and Activities for Driver Inattention Detection

Auto Companies

Many famous auto companies are currently conducting researches on driver inattention monitoring systems, including Toyota, Nissan, Volvo, Mercedes-Benz, and Saab.

Saab's Driver Attention Warning System [41, 42] is a project designed to counter two of the most common causes of road accidents: driver drowsiness and distraction. The system utilizes two miniature infrared cameras, one installed at the base of the driver's A-pillar and the other in the center of the main fascia, which are focused on the driver's eyes. It also utilizes SmartEye [43] software to get accurate eyelid, gaze, and head orientation information. In their algorithm, the driver's eye blinking frequency is measured. If a pattern of long duration eyelid closures is detected, it indicates the potential onset of drowsiness. A three-level warning interface was designed for drowsiness detection. This starts with a chime sound and text message, moves on to a spoken message, and finally a stronger warning tone audio message is persistently delivered until the driver presses the reset button. As soon as the driver's gaze moves away from what is defined as the "primary attention zone"-the central part of the windshield in front of the driver-a timer starts counting. If within 2 s of the timer being triggered the driver's eyes and head do not return to the "straight ahead" position, it is considered to be a distraction. In a case involving peripheral tasks such as looking in the rearview mirror, a side mirror, or turning a corner, the timer's elapse time becomes longer. Once the driver distraction has been detected, a seat vibration signal will be issued to warn the driver. However, there is no report about the robustness of this system during daytime and nighttime driving under different kinds of weather conditions, providing no driver status ground truth as a reference.

Toyota developed their Driver Monitoring System in 2006 for its latest Lexus models. This system features a camera, using near-infrared technology, mounted on top of the steering column cover. It monitors the exact position and angle of the driver's head while the vehicle is in motion. If the Advanced Pre-Crash Safety system detects an obstacle ahead, and at the same time the Driver Monitoring System establishes that the driver's head has been turned away from the road for too long, the system automatically activates pre-crash warnings. If the situation persists, the system can briefly apply the brakes to alert the driver [44]. In 2008, the Toyota Crown system went further. It can detect if the driver is becoming sleepy by monitoring their eyelids. Toyota's solution combines driver face orientation and environmental obstacle detection to determine accident potential, and utilizes eyelid activity to identify drowsiness.

In the spring of 2009, Mercedes-Benz introduced Attention Assist into its series production [45]. Attention Assist works by first observing a driver's behavior, and then using this information to create a unique driver profile. During operation, a series of tests continually monitor the driver input in relation to this profile, and in the event that a deviation is encountered, the system then determines whether or not the deviation is a result of fatigue. If it is, Attention Assist alerts the driver both visually and audibly that it is time to take a break. The factors taken into account to determine a driver's profile include the speed, longitudinal and lateral acceleration, angle of the steering wheel, way that the indicators and pedals are used, certain driver control actions, and even various external influences such as a side wind or an uneven road surface. The Attention Assist system only uses vehicle parameters to determine driver drowsiness, which requires no additional hardware setup. However, this system needs to establish individual profiles for different drivers, which would affect the acceptance of the system in real life.

In 2007, Volvo Cars introduced Driver Alert Control to alert tired and nonconcentrating drivers [46]. With the idea that the technology for monitoring a driver's eyes is not yet sufficiently mature and human behavior varies from one person to another, they developed their system based on the cars progress on the road. It is reported that Driver Alert Control

monitors the car's movements and assesses whether the vehicle is being driven in a controlled or uncontrolled way. It can also cover situations where the driver is focusing too much on his or her cell phone or children in the car, thereby not having full control of the vehicle.

Other Commercial Products and Activities

Technological approaches have continued to emerge in recent years, and hold promise for detecting and monitoring dangerous levels of driver inattention. While many of these projects are now in the development, validation testing, or early implementation stages, some companies can provide usable devices or prototype to give information about driver behavior. For the purpose of nonintrusive measurement, these devices mainly utilize video cameras and computer vision technologies.

Attention Technology, Inc. has designed and developed the DD850 Driver Fatigue Monitor (DFM), the only real-time, on-board drowsiness monitor that is currently being tested in an extensive field operational test. The DFM is a video-based drowsiness detection system that works by measuring slow eyelid closure. It is designed to mount on a vehicle's dashboard just to the right of the steering wheel, and provides a continuous real-time measurement of eye position and eyelid closure [47]. Specifically, the DFM estimates PERCLOS to determine drowsiness, which is the proportion of time the eyes are closed 80% or more over a specified time interval. The DFM uses a structured illumination approach to identify the driver's eyes. This approach obtains two consecutive images of the driver using a single camera. The first image is acquired using an infrared (IR) illumination source that produces a bright-pupil image. The second image uses an IR illumination source at a different wavelength to produce an image with dark pupils. These two images are essentially identical except for the brightness of the pupils in the images. A third image calculates the difference between these two images, enhancing the bright eyes and eliminating all image features except for the bright pupils. The driver's eyes are identified in this third image by applying a threshold to the pixel brightness. The bright pupil effect utilized by the DFM is a simple and effective eye-tracking approach for pupil detection based on a differential lighting scheme. However, the success of the bright pupil technique strongly depends on the brightness and size of the pupils, which are often functions of face orientation, external illumination interference, the distance of the subject from the camera, and race. For real-world in-vehicle applications, sunlight can interfere with IR illumination, reflections from eyeglasses can create confounding bright spots near the eyes, and sunglasses tend to disturb the IR light and make the bright-pupil phenomenon appear very weak.

Delphi believes that computer vision offers the most direct method for detecting the early onset of sleepiness and distraction, and it is also seen as an excellent platform to be shared with other vision-based driver assistance applications in the future. They integrated two products, the ForeWarn Drowsy Driver Alert and the ForeWarn Driver Distraction Alert, into a comprehensive Driver State Monitor (DSM) [47]. The DSM is a computer vision system that uses a single camera mounted on the dashboard directly in front of the driver and two IR illumination sources. Upon detecting and tracking the driver's facial features, the system analyzes eye closures and head pose over time to infer fatigue or distraction level. It provides an extended eye-closure warning for closures longer than 2.5 s, and also provides an extended distraction warning for nonforward gaze states in excess of 2.5 s. The fatigue detection algorithm predicts AVECLOS, the percentage of time the eyes are estimated to be fully closed over a 1-min interval. Because this is a less complex measure of drowsiness than PERCLOS, it permits the use of an automotive-grade data processor, in contrast to the high-grade PC processor required for PERCLOS.

Seeing Machines is engaged in the research, development, and production of advanced computer vision systems for research on human performance measurement, advanced driver assistance systems, transportation, etc. [48]. Their signature product, faceLAB, provides head and face tracking, as well as eye, eyelid, and gaze tracking for human subjects, using a noncontact, video-based sensor. faceLAB provides comprehensive blink analysis and PERCLOS assessment, including the delivery of raw data on the details of eyelid behavior. Instead of using traditional corneal reflection techniques, input is obtained using a stereo

camera pair. Seeing Machines's faceLAB has been employed extensively as a PC-based research tool. Although the device reportedly works very well in a simulator environment, the numerous challenges faced in a real driving environment prevent it from working robustly. Seeing Machines also provides another product, Driver State Sensor (DSS). It consists of one camera, two IR LED illuminators, and one special computing and communication unit. The goal of DSS is to detect driver fatigue by analyzing eyelid activity.

Smart Eye AB is another company that provides computer vision–based software for detecting human face/head movement, eye movement, and gaze direction [43]. Their product, Smart Eye Pro 3.0, is a machine vision system that estimates head pose using a simple and robust method based on tracking individual facial features and a three-dimensional (3D) head model. While the face is being tracked, the gaze direction and eyelid positions are determined by combining image edge information with 3D models of the eye and eyelids. A major advantage is that eye and head tracking can continue even if one camera is fully occluded or otherwise nonoperational. This also allows for large head motions (translation and rotation). Smart Eye has not developed an algorithm that monitors drowsiness.

SensoMotoric Instruments GmbH (SMI) [49] is a German company. Their product, InSight, can measure head position and orientation, gaze direction, eyelid opening, and pupil position and diameter. It uses a sampling rate of 120 Hz for head pose and gaze measurement, 120 Hz for eyelid closure and blink measurement, and 60 Hz for combined gaze, head pose, and eyelid measurement. It also provides PERCLOS information for drowsiness detection. It is a computer based system and needs user calibration.

Current Methods to Detect Driver Inattention

In the scientific literature, five main types of measures for inattention detection are commonly used: (a) subjective report measures, (b) driver biological measures, (c) driver physical measures, (d) driving performance measures, and (e) hybrid measures. With the exception of subjective report measures, these measures are based on nonlinear modeling techniques. This section briefly reviews the most common nonlinear modeling techniques. Then, the researches on the five main types of measures will be explored. Finally, the extraction of physical signals from a driver by image processing will be discussed at the end of this section, since driver physical measures offer distraction detection through eye gaze monitoring and fatigue detection through eye gaze, blink, head, and mouth tracking.

Nonlinear Modeling Techniques

Human cognition can hardly be represented by a linear model. Hence, nonlinear modeling techniques are greatly adopted in the driver inattention detection area. Nonlinear modeling with machine learning techniques can extract information from noisy data, and do not require prior knowledge before training. There are also some mechanisms in machine learning that can avoid over-fitting for nonlinear modeling, producing more robust and general models than traditional learning methods (e.g., logistic regression), which only minimize training error.

Artificial neural networks (ANNs) have been studied and utilized in numerous scientific and engineering fields. One of the main advantages of ANNs is that they infer solutions from data without any prior knowledge of the patterns in the data, that is to say, they extract the patterns empirically even if the equation between the inputs and outputs does not exist. This characteristic is very important because in most practical cases the exact input-output relationship is difficult to establish. ANNs also have the ability to generalize (i.e., they respond with a reasonable accuracy to patterns that are broadly similar to the original training patterns), which is very useful because real-world data is noisy, distorted, and often incomplete. ANNs are nonlinear, which allows them to solve some complex problems more accurately than linear techniques [27].

The fuzzy inference system (FIS) is famous for its well-known linguistic concept modeling ability. The fuzzy rule expression is close to an expert natural language. A fuzzy system then manages the uncertain knowledge and infers high-level behaviors from the observed data. On the other hand, as it is a universal approximator, the FIS can be used for knowledge induction processes [50].

Support vector machine (SVM) is based on the statistical learning technique and can be used for pattern classification and the inference of nonlinear relationships between variables. This method has been successfully applied to the detection, verification, and recognition of faces, objects, handwritten characters and digits, text, speech, and speakers, along with the retrieval of information and images [51]. The learning technique of the SVM method makes it suitable for measuring the cognitive states of humans. SVMs can generate both linear and nonlinear models and are able to compute the nonlinear models as efficiently as the linear ones. Given a set of input data, this method first transforms the input domain through a kernel, and then looks for a hyperplane in the transformed domain that separates the data with minimum error and maximum gain. Finally, the hyperplane is transformed back to the input domain to obtain the decision boundaries, which may potentially be nonlinear.

AdaBoost is a learning algorithm that uses the pattern recognition algorithm called Boosting [52]. Its advantages include a high classification performance, fast recognition process time, and the potential extension of recognition features. In AdaBoost, learning involves the creation of different classifiers while successively changing the weighting of the learning data. A weighted majority decision is then made for these multiple classifiers in order to obtain the final classifier function. Individual classifiers are referred to as "weak classifiers," while the combination of classifiers is a "strong classifier."

Bayesian networks (BNs) have several advantages that make them well suited for describing human behavior. First, the hierarchical structure of BNs can systematically present information from different sources and at different levels of abstraction, and can also capture probabilistic relationships. Second, a BN is not only a computational model but also a form of knowledge representation. Unlike other data mining approaches such as the SVM, BNs reveal the relationships that generate the model predictions. Third, BNs can handle situations with missing data. The certainty of the hypothesis will change according to the BN's reasoning, which incorporates new data using a probabilistic dependence network when new evidence is added. Because of these advantages, BNs are applicable to human-behavior modeling and have been used to detect inattention [53]. Despite these advantages, creating a correct and stable BN model requires extensive computational capability and a large amount of training data.

Another emerging trend has been to borrow techniques based on hidden Markov models (HMMs) from the speech processing and language technology field and apply these to driver behavior modeling for route recognition, driver identification, and distraction detection in a manner analogous to speech recognition, speaker identification, and stress detection in speech [54]. The foundation of HMM is a stochastic Markov process consisting of a number of states with corresponding transitions. At discrete time intervals, the Markov process moves from one state to another according to a set of transition probabilities. State changes in the Markov process are hidden from the user. Sathyanarayana et al. [54] constructed a hidden Markov model using vehicle speed, steering wheel angle, and braking force to predict route maneuvers (left turn, right turn, and lane change).

Subjective Report Measures

The Karolinska sleepiness scale (KSS) is the most commonly used tool for the subjective self-assessment of sleepiness; the values used in the KSS are shown in Table 1. Kaida et al. [55] investigated the validity and reliability of the KSS using EEG, behavioral, and other

Driver Inattention Monitoring System for Intelligent Vehicles. Table 1 Karolinska sleepiness scale (KSS)

KSS	Meaning
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, no effort to stay awake
8	Sleepy, some effort to stay awake
9	Very sleepy, great effort to keep awake, fighting sleep

subjective indicators of sleepiness. Their study showed that the KSS was closely related to EEG and behavioral variables, which indicates that the KSS has a high validity for measuring sleepiness.

Ingre et al. [56] verified the close relationship between subjective sleepiness measured with the KSS and blink duration (BLINKD) and lane drifting, calculated as the standard deviation of the lateral position (SDLP) in a high-fidelity moving base driving simulator. Their experiments showed a significant effect of the KSS on both BLINKD and SDLP. A test for a quadratic trend suggested a curvilinear effect with a steeper increase at high KSS levels for both SDLP and BLINKD. Craig et al. [57] also found that psychological factors correlated consistently with self-reported fatigue. However, the KSS is recorded over relatively long time intervals, say every 15 min, as a trade-off between high temporal resolution and avoiding intrusive feedback. As a consequence, the KSS was not capable of recording sudden drowsiness variations caused by different situations.

Driver Biological Measures

Biological signals include EEG, electrocardiogram (ECG), electro-oculography (EOG), surface electromyogram (sEMG), etc. These signals are collected through electrodes in contact with the skin of the human body. Table 2 summarizes some typical methods used in this field. EEG has a spatial resolution of 20 mm and a temporal resolution of 0.001 s. It is widely used in the brain activity research field. Recent research has proposed various methods to extract features from a segment of raw EEG data for fatigue detection. In the time domain, the average value, standard deviation, and sum of the squares of EEG amplitude are the most commonly used features. In the frequency domain, the energy content of each band (β, α, θ, δ),

Driver Inattention Monitoring System for Intelligent Vehicles. Table 2 Summary of biological measures

References	Bio-signal type	Objective	Analysis methods	Result
[58]	EEG	Fatigue	Assess δ, θ, α, β	$(\theta + \alpha)/\beta \uparrow$
[59]			Sample entropy, phase synchronization	Sample entropy, phase synchronization $\uparrow$
[60]			SVM	Predict alert $\rightarrow$ drowsy
[61]			Probabilistic-SVM	Better than standard SVM
[62, 63]			ICA, FFT, correlation analysis, LRM	Est. drowsy level with 87% accuracy
[64]			KPCA algorithm	Complexity decreases as fatigue increases
[23]		Mental engagement	Inspection with second timescale	Different mental task can be detected
[65]			XCS	Different mental task can be detected
[66]	EEG, ECG	Fatigue	Dynamic Bayesian network	More features are favorable
[67]	ECG, PPG	PVT	Multi linear regression model	ECG, PPG is useful for est. PVT
[68]	EOG	Hypovigilance	Eight eye actives, fuzzy expert sys	Pre. sleep-related accidents with high acc
[69]	EOG	Drowsiness	Eleven eye actives, SVM	Accuracy is quite high for "very sleepy"
[70]	sEMG	Fatigue	Statistic analysis	Statistic trends were given
[71]			Frequency and statistic analysis	MDF, MNF, RMS show large change

mean frequency, and center of gravity of the EEG spectrum are commonly used. Other models such as the auto-regressive moving average (ARMA) and power spectrum estimation are also used by some researchers to extract EEG features. The most reliable patterns in terms of their consistency and occurrence for fatigue are the β, α, θ, and δ waves (see Fig. 3).

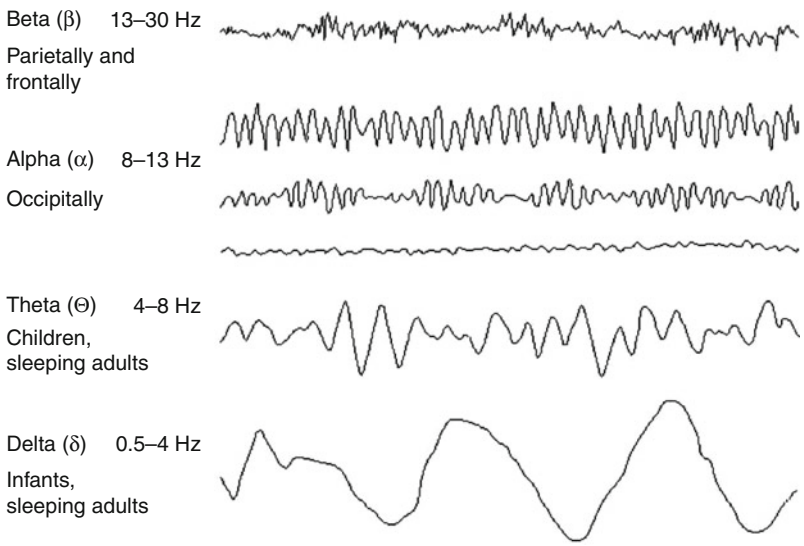
EEG is widely accepted as a good indicator of the transition between wakefulness and sleep, as well as between the different sleep stages. It is often referred to as the gold standard. Svensson [29] proposed objective sleepiness scoring (OSS), derived from EEG signals, as the ground truth for validating other drowsiness detection algorithms. The five-level OSS scores are described in [29], and are shown in Table 3.

The four EEG activities (β, α, θ, and δ) were assessed in [58] for 52 subjects during a monotonous driving session. The results showed an increasing trend for the ratio of slow wave to fast wave EEG activities over time. In [59], sample entropy and phase synchronization were adopted to detect fatigue from EEG signals, with the results showing that phase synchronization among the hemispheres gradually increased and sample entropy decreased, both pointing to a gradual increase in sleepiness, which is related to a decrease in EEG complexity. Yeo et al. [60] trained the SVM to classify EEG signals into four principal frequency bands, and then to predict the transition from alertness to drowsiness. Shen et al. [61] compared a probabilistic-based multi-class SVM and standard multi-class SVM as classifiers for distinguishing mental

Driver Inattention Monitoring System for Intelligent Vehicles. Table 3 Objective sleepiness scoring (OSS)

OSS score	EEG content
0	Background of continuous β waves, no α, no θ waves
1	Occurrence of α and/or θ waves, in at least two regions of the brain, for less than a cumulative length of 5 s
2	Occurrence of α and/or θ waves, in at least two regions of the brain, for less than a cumulative length of 5 s
	or
	Occurrence of α and/or θ waves, in at least two regions of the brain, for less than a cumulative length of 5 s
3	Occurrence of α and/or θ waves, in at least two regions of the brain, for less than a cumulative length of 5 s
4	Continuous α and/or θ waves

Source: Derived from EEG data [29]



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 3
Four types of EEG waves [29]

fatigue into five mental-fatigue levels, and showed that the accuracy of the probabilistic-based multi-class SVM was better. Lin et al. [62] established a linear regression model to estimate the drowsiness level from the ICA of 33-channel EEG signals, and could estimate the drowsiness level with 87% accuracy. They then implemented a real-time embedded EEG-based driver drowsiness estimate system in [63], which adopted only four channels of EEG data.

Although not apparently related, some researchers have shown that it is also possible to estimate the distraction level from EEG data. Berka et al. [23] tried to use EEG data to continuously and unobtrusively monitor the levels of task engagement and mental workload in an operational environment. An inspection on the EEG data using a second-by-second timescale revealed associations between the workload and engagement levels when aligned with specific task events, which provided preliminary evidence that second-by-second classifications reflect parameters of task performance. Skinner et al. [65] investigated the efficacy of the genetic-based learning classifier system XCS in classifying artifact-inclusive EEG signals into four mental tasks designed to elicit hemispheric responses. In Liu et al. [64], the KPCA algorithm was employed to extract nonlinear features from the complexity parameters of EEG (approximate entropy (ApEn) and Kolmogorov complexity (Kc)) and improve the generalization performance of an HMM. The result showed that both complexity parameters decreased significantly as the mental fatigue level increased, and the classification accuracy reached 84%.

In addition to EEG, other biological signals such as ECG, EOG, and sEMG have also been tested to estimate driver mental state. Yang et al. [66] employed a dynamic Bayesian network with EEG and ECG to estimate fatigue. A first-order hidden Markov model (HMM) was employed to compute the dynamics of a Bayesian network at two different time slices. The results showed that more features are favorable for inferring the driver fatigue more reliably and accurately. In Chua et al. [67], a multiple linear regression model was established to estimate PVT (psychomotor vigilance test) values from a combination of ECG and photoplethysmogram (PPG) data. Damousis and Tzovaras [68] selected eight eye activity features, extracted from EOG, to develop a fuzzy expert system (FES) for the detection of hypovigilance. Shuyan and Gangtie [69] employed an SVM to perform drowsiness prediction with 11 eyelid-related features extracted from EOG. These eyelid features included blink duration, blink duration 50–50, amplitude, lid closure speed, peak closing velocity, lid opening speed, peak opening velocity, delay of eyelid reopening, duration at 80%, closing time, and opening time. It was reported that the drowsiness detection accuracy was 86% for "sleepy." In Balasubramanian and Adalarasu [70], the surface sEMG of the shoulder and neck was analyzed while the participant was driving to determine the onset of fatigue and prove that the development of muscular fatigue is a consequence of driving. In Katsis et al. [71], frequency analysis and statistical analysis were performed on sEMG signals from the left bicep, right bicep, left forearm flexor, right forearm flexor, and frontal muscles. The results showed that the middle frequency decreased by about 9.5–18.9%, the mean frequency decreased by about 11.3–18.4%, and the root mean square amplitude increased by about 25.1–47.7% from their initial values for a predefined driving route.

Driver Physical Measures

Fatigue Detection In Bergasa et al. [50], PERCLOS, eye closure duration, blink frequency, nodding frequency, fixed gaze, and frontal face pose were normalized and used as inputs to the fuzzy inference system for fatigue detection. Different linguistic terms and their corresponding fuzzy sets were distributed in each of the inputs using induced knowledge based on the hierarchical fuzzy partitioning (HFP) method. They then chose the fast prototyping algorithm with the pruned method (FDT + P) to automatically generate fuzzy rules that were consistent, lacked redundancy, and were interpretable. Afterward, a simplification process was applied to achieve a more compact knowledge base to improve the interpretability and maintain the accuracy. Finally, three variables (fixed gaze, PERCLOS, and eye closure duration) were determined to be crucial cues for detecting a driver's fatigue. By fusing them with a fuzzy system, a final fatigue detection accuracy of 98% was achieved.

Fan et al. [72] utilized a Gabor-features representation of the face for fatigue detection. After the face was

located, Gabor wavelets were applied to the face area to obtain different scale and orientation features of the face. Then, features on the same scale were fused into a single one to reduce the dimension. Finally, the AdaBoost algorithm was used to extract the most critical features from the dynamic feature set and construct a strong classifier for fatigue detection. It was reported that this method worked well on a wide range of human subjects with different genders, poses, and illuminations.

Friedrichs and Yang [73] explore 18 features of eye movement for drowsiness detect. The features are listed in Table 4. Rather than using principal component analysis (PCA) or linear discriminate analysis (LDA) to reduce the dimension of the features, they chose the sequential floating forward selection (SFFS) [74] algorithm to select the most promising features to construct a classifier. The advantage of SFFS over feature transform techniques is its high transparency, because the selected features remain unchanged. An ANN classifier was trained to detect the drowsiness, and the results showed that as long as the blinking signals were correctly detected (high confidence), the drowsiness detection accuracy could reach 82.5%.

Driver Inattention Monitoring System for Intelligent Vehicles. Table 4 Eighteen features of eye movement [73]

Features
Average eye closure speed
Amplitude/velocity ratio (APVC)
APCV with regression
Blink amplitude
Blink duration
BLINKDUR baselined
Blinking frequency
Energy of blinking (EC)
EC baselined
Microsleep event 0.5 s rate
Microsleep event 1.0 s rate
Mean square eye closure
Mean eye closure
Percentage eyes >70% close (PERCLOS70)
Percentage eyes >80% closed (PERCLOS80)
PERCLOS70 baselined
PERCLOS80 EWMA baselined
Head nodding

Some other methods have also been used for fatigue detection. In Sun et al. [75], a Bayesian network was employed to infer fatigue from gaze information. Orazio et al. [76] used a mixture Gaussian model to model the "normal behavior" statistics from the eye closure duration (ECD) and frequency of eye closure (FEC) for each person, in order to identify anomalous behaviors. Suzuki et al. [77] derived three factors from the blinking waveform: the length of a blink, the closure rate, and the blink rate. These factors were then weighted using a multiple regression analysis for each individual to calculate the drowsiness level. In Senaratne et al. [78], four cues were fused using fuzzy logic to detect driver fatigue: PERCLOS, head nodding frequency, slouching frequency, and posture adjustment frequency. In addition to analyzing eye activities, some researches also analyzed mouth activities [79–81] to estimate the level of driver inattention. Fan et al. [80] used an LDA to classify the mouth into two states: normal and yawning. In Vural et al. [81], they used a BP ANN to estimate three mouth states from lip features: normal, yawning, and talking. Vural et al. [81] used a facial action coding system (FACS) to code facial expressions, and then employed machine learning to discover which facial configurations were suitable for fatigue detection, with 31 facial actions employed to predict drowsiness. This system claimed to be able to predict sleep-and-crash episodes with a 96% accuracy within subjects, and an accuracy above 90% across subjects.

Distraction Detection Kircher et al. [82] described and compared two different algorithms for gaze-based driver distraction detection, based on the eye-tracking data obtained in a field study. One algorithm relied on the metric "percent road center" (PRC) of gaze direction, where a PRC of more than 92% was considered to be indicative of a gaze concentration resulting from cognitive distraction, while a PRC below 58%, computed over 1 min, was a indicator of visual distraction. Fixations were used for the computation of PRC. The second algorithm was based on a 3D world model

with different interior zones such as the windshield, speedometer, mirrors, dashboard, etc., and on the time the driver spends glancing at those zones. A time-based "attention buffer" with a maximum value of 2 s was decreased over time when the driver looked away from the "field relevant for driving" (FRD), while it was increased when the driver's glance was inside the FRD, until the maximum value was reached. When the buffer reached zero the driver was considered to be distracted, and when further conditions were met (direction indicator not activated, speed above 50 km/h, no brake activation, and no extreme steering maneuvers), a warning was issued. The results showed that both algorithms have potential for detecting driver distraction, and that fully attentive drivers had a PRC of about 70–80%.

Pohl et al. [83] used head pose and eye gaze information to model the visual distraction level, which was time dependent on the visual focus, with the assumption that the visual distraction level was nonlinear: Visual distraction increased with time (the driver looked away from the road scene) but decreased nearly instantaneously (the driver refocused on the road scene). Based on the pose/eye signals, they established their algorithm for visual distraction detection. First, they used a distraction calculation to compute the instantaneous distraction level. Then, a distraction decision-maker determined whether the current distraction level represented a potentially distracted driver.

Bergasa et al. [84] tried to detect visual distraction with head pose and fatigue with yawning, eyebrow raising, and PERCLOS. Although they developed an algorithm for extracting the required cues, the algorithm for fusing them was unclear.

Driving Performance Measures

A change in the mental state can induce a change in driving performance. In Furugori et al. [85], the pressure distribution on the seat of male subjects was measured during simulated long-term driving, and the results showed there was a relationship between changes in the load center position (LCP) and driver reported subjective fatigue. Their algorithm to derive a fatigue index was calculated on a time interval of 10 min, which was a considerable delay.

Farid et al. [86] tried to distinguish between attentive and inattentive driving in car-following situations by analyzing the vehicle following distance and steering angle. They built up a real-time model using hidden Markov models with Gaussian mixtures to infer the intentions of the driver, and this model was able to detect a lane change half a second earlier than conventional approaches. Zhong et al. [36] performed a localized energy analysis of the steering wheel angle dynamics and vehicle tracking to detect driver fatigue, and found a trend of localized energy increase with driving time. In Takei and Furukawa [87], chaos theory was employed to explain the dynamics of steering wheel motion, and estimate driver fatigue. Using a proper time delay, they found the attractors, which involved the Chaos characteristics. They stated that they will study the Lyapunov exponent of this chaos to estimate the driver fatigue. In addition to an energy analysis, in [88], a Gaussian mixture model was adopted to identify the driver based on the driving behavior signals: forces on the pedals and vehicle velocity.

Kari Torkkola and Wood [89] adopted the steering wheel position, accelerator pedal position, lane boundaries, and upcoming road curvature to infer driver status. First, the original signals were preprocessed (averaging, entropy, etc.), which yielded a huge set of features. Then, random forest (RF) [90], a technique based on ensembles of learners, was employed to select the optimal parameters from the derived features. The classifier was also constructed using RF, and the final accuracy reached 80%.

In Ersal et al. [91], a radial-basis neural-network-based modeling framework was developed to characterize normal driving behavior. Then, in conjunction with an SVM, it was able to classify normal and distracted driving. Vehicle dynamics and driving performance data such as vehicle position, velocity, and acceleration, as well as throttle and brake pedal positions, were adopted to model normal driving. The average and standard deviations of the residuals (the differences between the actual and model-predicted driver actions) were chosen as the inputs for the SVM. The results showed that the accuracy varied between individuals.

Hybrid

Combining driver physical measures with driving performance measures could intuitively increase the inattention detection confidence. On the other hand,

road scene analysis and observations of the driver's face would make it possible to estimate what the driver knows, what the driver needs to know, and when the driver should know it. Combining driver gaze information with road scene information offers several potential benefits: context-relevant information selection, unnecessary information suppression, and anticipatory information selection. Table 5 shows a summary of some researches that utilized a hybrid method for detecting driver inattention.

Fatigue Detection Eskandarian et al. [27] utilized artificial neural networks (ANNs) to analyze vehicle parameter data and eye closure data to infer driver fatigue. The vehicle parameter data included speed, acceleration, vehicle lane position, steering angle, braking, and heading angle, which was recorded at a frequency of 20 Hz. The eye closure data was recorded at 60 Hz using PC-based equipment by Applied Science Laboratory (ASL), which recorded pupil diameter by capturing reflections from the pupils (bright pupil). Then they analyzed the data to identify the potential variables that were correlated with drowsiness. This analysis found four variables highly correlated with fatigue: PERCLOS, vehicle crash, vehicle lateral displacement, and steering wheel angle. Out of consideration for the simplicity and robustness of the data acquisition, Eskandarian et al. [27] implemented two ANN fatigue detectors: one utilized the steering wheel angle signal as an input and the other utilized both the steering wheel angle signal and the eyelid signal as inputs. The steering angle was preprocessed before being input in the ANN. The preprocessing scheme involved normalization for road curvature, discretization at different ranges, coding, and 15-s accumulation. For the eyelid signal, the preprocessing scheme was the same as that for the steering angle, except for eliminating the normalization stage. It was reported that after proper training and cross validation, the steering-eye ANN had an accuracy of 88%, with

Driver Inattention Monitoring System for Intelligent Vehicles. Table 5 Summary of hybrid measures

References	Raw signals	Fusion technique	Object
[27]	Vehicle parameter data and eye closure data	ANN	Fatigue detection
[18]	Eye gaze, head orientation, diameter of pupils, heart rate (RRI)	SVM and Adaboost	Cognitive distraction detection
[92]	Leg and head motions, CAN signals	K-nearest neighbors classifier	Distraction detection
[93]	Audio signal and CAN signals	GMM/UBM	Distraction detection
[94]	Head orientation and the surround salience map	Direct matching	Visual distraction detection
[95]	Gaze variables, driving data and road geometry	ANOVA and binary logistic regression.	Distraction detection
[51]	Eye movement and vehicle parameters	SVM	Cognitive distraction detection
[53]	Eye movement and vehicle parameters	Bayesian network	Cognitive distraction detection
[96]	Head/eye and vehicle parameters	SVM	Visual and cognitive distraction detection
[97]	Vehicle and environment parameters	ANFIS	Distraction estimation
[98]	Eye gaze, blink, head pose, and environment parameters	Region matching	Visual and cognitive distraction detection
[99]	Head dynamics, facial features, upper body posture information and vehicle dynamics	In developing	Driver assistance

a false alarm rate of 9%, while the steering-only ANN had an accuracy of 85%, with a false alarm rate of 14%.

Distraction Detection In Miyaji et al. [18], the standard deviations of eye gaze, head orientation, pupil diameter, and average heart rate (RRI) were combined to improve the accuracy of the driver cognitive distraction detection. The eye and head parameters were obtained using faceLAB, while the RRI data came from ECG. In Miyaji et al. [18], two machine learning techniques, SVM and Adaboost, were implemented under the same conditions. The results showed that the classification performance of Adaboost was slightly better than that of SVM, while the recognition time of AdaBoost was approximately 1/26 that of the SVM.

Sathyanarayana et al. [92] tried to detect distraction by combining motion signals from the leg and head with vehicle signals. The motion signals included the three-axis acceleration of the right leg and two-axis orientation of the head. The vehicle signals adopted included vehicle speed, braking, acceleration, and steering angle. Then, a group of features were derived from these signals based on the nature of the signals. The feature types are listed in Table 6. Next, these derived features were analyzed using LDA to reduce the dimension. Then, a K-nearest neighbors classifier was trained and verified.

Driver Inattention Monitoring System for Intelligent Vehicles. Table 6 Candidate signal features [92]

Signal features
Maximum value of the signal
Minimum value of the signal
Amplitude of the difference between the first value and the last value
Duration of the signal
Maximum difference between any two consecutive values
Median of the signal
Mean of the signal
Difference between the maximum and minimum value of the signal
Standard deviation of the signal
Root mean square value of the signal
Difference between the maximum and minimum value of the differential of signal

In order to cope with the variability between drivers and maneuvers (context), [93] utilized a GMM/universal background model (UBM) and likelihood maximization learning scheme to first identify the driver through an audio signal and then recognize the maneuvers (right/left turn and lane change) through CAN-bus signals. Finally, the CAN-bus signals were also used to detect distraction for a particular driver and particular maneuver. It was reported that this system could reach an accuracy of 70% for distraction detection.

Doshi and Trivedi [94] fused head orientation detection and a saliency map of the surroundings to determine whether there was a salient object in the driver's view, which gave an indication of whether a driver's head turn was motivated by the goal in his or her mind or some distracting object/event in the environment.

It is known that road geometry influences gaze behavior [100], and this aspect was taken into account by including road geometry as an additional factor when detecting driver distraction in [95]. They utilized an analysis of variance (ANOVA) and binary logistic regression to analyze and establish a model for distraction detection based on gaze variables and driving data: fixations (number and duration), scan path, standard deviation of gaze location, speed (minimum, maximum, average and percentage change in speed), lateral acceleration (maximum), and longitudinal deceleration (maximum). The results showed that the road geometry does influence the accuracy of distraction detection based on driving data, but gaze behavior is mainly influenced by distraction, with little or no influence by road geometry.

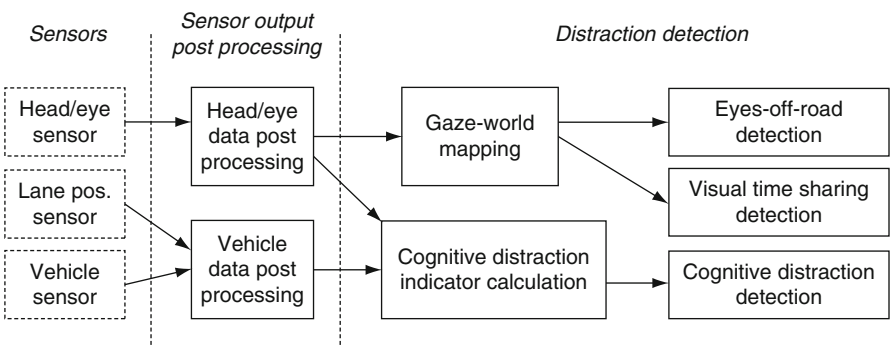
Liang et al. [51] tried to detect the driver cognitive distraction caused by interacting with In-Vehicle Information Systems (IVISs) in real time by fusing eye movement and driving performance using an SVM. The measured signals included fixation, saccade, smooth pursuit of eye (calculated from raw gaze vector obtained using faceLAB [48]), steering wheel angle, lane position, and steering error. These measures were summarized over various windows to create instances that became the SVM model inputs. After training, the SVM model could detect driver distraction with an average accuracy of 81.1% (sd = 9.05%). Lee et al. [53]

utilized the same conditions as [51] but adopted a Bayesian network to detect cognitive distraction, showing that compared to an SVM model, the dynamic Bayesian network produced better accuracy.

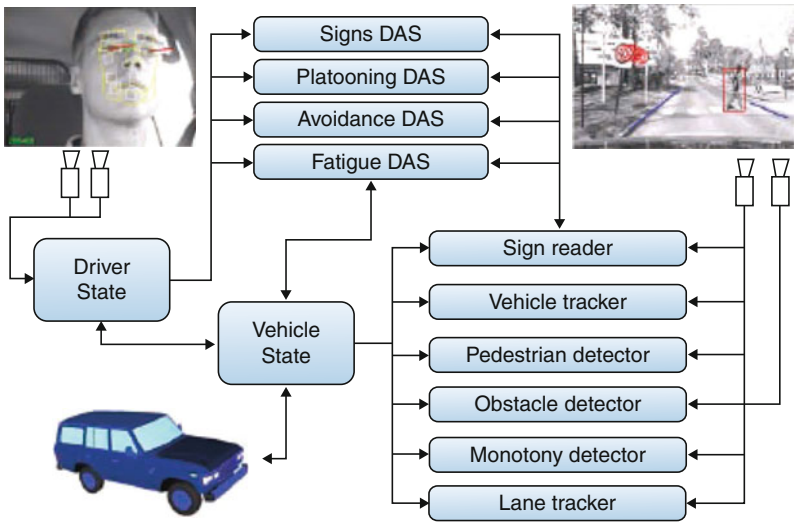
Markkula and Kutila [96] concentrated on processing head/eye and vehicle performance information to estimate both visual and cognitive distractions; their algorithm is shown in Fig. 4. The head/eye information derived from stereo cameras included head position, head orientation, gaze orientation, saccade, and blink identification, as well as confidence values. The vehicle performance information included lane position, vehicle speed, etc. Based on the head/eye information, they developed Gaze-World Mapping and Eyes-Off-Road Detection, which could detect momentary visual distraction. Another algorithm, visual time-sharing detection, was developed to measure longer term visual distractions. For cognitive distraction, they used three indicators to classify the cognitive tasks with an SVM: the standard deviation of gaze angle, standard deviation of head angle, and standard deviation of lane position. However, in the Gaze-World Mapping phase, which mapped gaze and head angles onto actual real-world targets of visual attention, the road-ahead target was static and determined off-line by inspecting the distribution of gaze angles for road-ahead data, and then manually enclosing the distribution in a rectangle.

Tango et al. [97] proposed a method to derive the distraction level from relevant vehicle and environment data using the adaptive neuro-fuzzy inference system (ANFIS). Rather than a binary "yes" or "no," they chose reaction time as the output to train, validate, and test their ANFIS model. The candidates to be selected as input for the ANFIS included the environment visibility, traffic density, and the standard deviations in speed, steering angle, lateral position, lateral acceleration, and deceleration jerk. After preprocessing, the level of difficulty of an IVIS and the standard deviation of steering angle were found to have the highest correlations with the reaction time. Thus, they were selected as the input. No accuracy information was provided in [97].

Fletcher and Zelinsky [98] utilized faceLAB to obtain information such as eye gaze direction, eye closure, and blink detection, as well as head position information. In this system, upper and lower bounds were placed on the percentage of time the driver spent observing the road ahead, called the percentage road center (PRC). A percentage that was too high (>90%) could indicate a fatigued state (e.g., vacant staring). A percentage that was too low (<20%) might indicate a distracted state (e.g., tuning radio). Similar to the PRC metric, they analyzed driver gaze to detect even shorter periods of driver distraction. They used gaze direction to reset a counter. When the driver looked forward at the road scene, the counter was reset. If the driver's gaze diverged, the counter began timing. When the gaze had been diverted for more than a specified time period, a warning was given. The time period for the permitted distraction was a function of the vehicle velocity. As the speed increased, the permitted time period would decrease, either as the inverse (reflecting time to impact) or the inverse squared (reflecting the



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 4
Overview of the distraction detection algorithms [96]



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 5
The distributed modular software architecture [98]

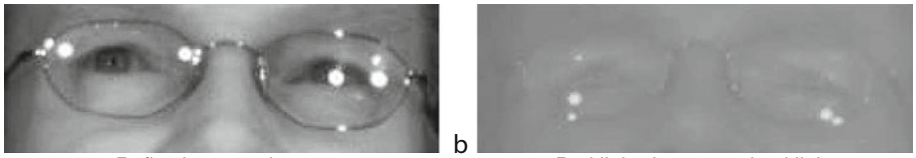
stopping distance). They tried to integrate driver gaze information into other driver assist systems to make the system more acceptable and safer. The framework is shown in Fig. 5. They also spent a significant amount of effort on integrating driver gaze information into lane tracking and sign reading systems. The lane-tracking system was used to orient the driver gaze information. A strong correlation was found to exist between the eye gaze direction and the curvature of the road during normal driving [101], with a slight correlation being a potential indicator of inattention. Fletcher and Zelinsky [98] integrated driver visual information with sign detection to implement a Sign Driver Assist System. This system recognized critical signs in the environment. At the same time, the driver monitoring system verified whether the driver looked in the direction of the sign. If it appeared that the driver was aware of the sign, the information could be made available passively to the driver. In contrast, if it appeared that the driver was unaware of the information, it could be highlighted.

A driver's body posture information is potentially related to driver intent, driver affective state, and driver distraction. Tran and Trivedi [99] explored the role of 3D driver posture dynamics in relation to other contextual information (e.g., head dynamics, facial features, and vehicle dynamics) for driver assistance. It focused on head pose and upper body posture extraction, but no significant results on driver assistance were found.

Driver Physical Signal Extraction

Numerous researches have adopted commercial eye trackers to obtain the physical signals related to the face/eye, which have allowed them to concentrate on exploring the inattention detection algorithm rather than image processing. However, these commercial eye trackers can only work well under specific constrained environments. They do not normally work well for real road conditions. For example, [73] adopted the driver state sensor from the seeing machines [48] to obtain the eye signal. However, even after many improvements, there were still some issues: Reflections from glasses led to bad signal quality and varying light conditions during daytime driving posed problems for the eye signal tracking (see Fig. 6). Therefore, much research has been conducted to make the physical signals extracted using image processing more accurate and robust. The methods for physical signal extraction are summarized in Table 7.

For face segmentation in driver inattention detection, the commonly used methods in the literature include a boosted cascade of Haar-like features [102], adaptive boosting [75], landmark model matching [78], skin color [79], and gravity center template [80].



Driver Inattention Monitoring System for Intelligent Vehicles. Figure 6
Image processing problems [73]

Driver Inattention Monitoring System for Intelligent Vehicles. Table 7 Summary of physical signal extraction

Face segment	Eye segment	Blink feature	Gaze	Mouth segment	Mouth feature	Head pose
Boosted cascade of Haar-like features [102], adaptive boosting [75], landmark model matching [78], skin color [79], gravity center template [80], disparity map [103]	Darkest regions search [102], template matching method [75], neural network [77], edge map [78], p-tile and k-means algorithm [104], Hough transform + neural classifier [76], "bright pupil" [50] [105, 106], Sobel edges + SVM [107], templates matching [108]	Optical flow [102], derivative [77], SVM [78], finite state machine (FSM) [50], Kalman filtering [105]	Hough transform and gradient direction [75], Kalman filters and the FSM [50], relative position between pupil and glint [105], pupil features + eigenspace [105], headband + 3D eyeball model [106]	Fisher classifier [79], gravity center template [80]	Connected component analysis to determine lip [79], Gabor wavelet to get mouth corners [80]	Two eyes and the center of face [104], pupil and the nostril position [50], headband [106]

Eren et al. [103] adopted stereo cameras and extracted a face from a disparity map on the assumption that the driver's face had a smaller depth than the background. They then used an embedded HMM to recognize the forehead, eyes, nose, mouth, and chin.

After the face area has been segmented, it is necessary to extract the eye area and mouth area for further processing to obtain physical signals. In the literature, the following methods have been employed to extract the eye area. In Brandt et al. [102], the eyes were extracted by assuming that they were the darkest regions in the face; Sun et al. [75] located the eyes using a template matching method; Suzuki et al. [77] used a neural network to detect the eyes; Senaratne et al. [78] used an edge map to locate the irises of the eyes; and Su et al. [104] used a p-tile algorithm and k-means algorithm to locate the eyes. In [76], candidate eye regions were first extracted using a modified Hough transform, then symmetric regions in the candidates were chosen as further candidates, and finally a neural classifier was used to validate the presence of the eye in the image.

Another popularly adopted method for locating eyes involves the use of the "bright pupil" effect produced by near-infrared light. In [50, 105, 106], a camera equipped with a two-ring IR illuminator was adopted to acquire a driver image. The ring sizes were empirically calculated to obtain a dark pupil image when the outer ring was turned on, and a bright pupil image when the inner ring was turned on. A controller was designed to synchronize the IR illuminator with the image frame rate: to ensure that

the images with and without bright pupils were interlaced. Digitally subtracting the dark pupil image from the bright pupil image produced a difference image, where the pupils appear to be the brightest regions in the image, as can be seen in Fig. 7. The pupils were detected on the resulting image by searching the entire image to locate two bright blobs that satisfied certain constraints. After locating the eyes in the initial frames, Bergasa et al. [50] used two Kalman filters, one for each pupil, to continuously and robustly monitor a driver with eye closure or oblique face orientation. Huang et al. [107] eliminated the need for the synchronizer by acquiring the pupil location from a single image: First, pupil candidates were obtained through Sobel edges, and then they were identified using an SVM with a Gaussian kernel. In Zhu et al. [108], a round template two values matching algorithm was proposed for locating bright pupils, which had an accuracy of 96.4% but consumed 1,011 ms on a PIII 800 MHz computer.

After the location of the eye is extracted, the blinking and gaze parameters should be calculated. In Brandt et al. [102], blinks were measured by analyzing the optical flow of the eye region. Suzuki et al. [77] used a derivative method to detect the eyelids and produce a blinking waveform. Senaratne et al. [78] used an SVM to classify the state of the eye as open or closed to get the PERCLOS value. In the "bright pupil" condition, Bergasa et al. [50] implemented a finite state machine (FSM), with five states defined: tracking_ok, closing, closed, opening, and tracking_lost. The transitions between states were achieved from frame to frame as a function of the width/height ratio of the pupils. The ocular parameters such as eye closure duration, PERCLOS, eye closure/opening speed, and blink frequency were calculated as functions of the FSM. Ji and Yang [105] used Kalman filtering to track eyelid movements and compute the PERCLOS and average eye closure speed (AECS).

Driver Inattention Monitoring System for Intelligent Vehicles. Figure 7
Fields captured and subtraction. (**a**) Image obtained with inner IR ring. (**b**) Image obtained with outer IR ring. (**c**) Difference image [50]

The gaze was estimated by combining the Hough transform and gradient direction in [75], while Bergasa et al. [50] calculated the gaze based on the position and speed data using Kalman filters and the FSM. Ji and Yang [105] estimated gaze direction using information about the head movement and relative position between pupil and glint, with the gaze direction quantized into nine zones: left, front, right, up, down, upper left, upper right, lower left, and lower right. Cudalbu et al. [106] utilized a headband and a simplified 3D eyeball model to estimate the gaze orientation with an accuracy that varied from 1° to 3°.

Besides the eye, estimating the position of the mouth is also useful in fatigue detection. Rongben et al. [79] used a fisher classifier to extract the mouth area from the face region, while Fan et al. [80] used a gravity center template to extract the mouth area. Then, Rongben et al. [79] used connected component analysis to find the lips and [80] used a Gabor wavelet to get the corners of the mouth.

The head nodding frequency, slouching frequency, and posture adjustment frequency were derived from changes in the head position in [78]. Su et al. [104] clustered facial orientations into five clusters: frontal, left, right, up, and down, depending on the position of the eyes and the center of the face. Similarly, based on the pupil and nostril positions, Bergasa et al. [50] made a coarse 3-D face pose estimation. Ji and Yang [105] used an eigenspace algorithm to map seven pupil features (inter-pupil distance, sizes of left and right pupils, intensities of left and right pupils, and ellipse ratios of left and right pupils) to determine face orientation, which was quantized into seven angles: −45°, −30°, −15°, 0°, 15°, 30°, and 45°. Cudalbu et al. [106] employed a headband with IR reflective markers to estimate the 6 DOF head pose with an average error of 0.2°.

Discussion: Future Directions

Issues with Detection

The subjective report measures can produce some reasonable results in quantifying fatigue level. Because this kind of approach requires that the driver report his or her state frequently, both the fatigue level result and the driver himself or herself could cause interference. Large individual differences have been observed with the overall driving performance and blink duration independent of the KSS values [56]. In addition, [109] demonstrated that drivers have difficulty judging their fitness, especially after about 3 h of continuous monotonous daytime driving with increasing drowsiness. For these reasons, it is not sufficient to solely record the KSS. However, if only a rough fatigue level is needed and the lowest cost is required, this kind of approach may be the best choice. The driver biological measures directly measure biological signals from a driver's body, and have been found to be highly accurate when used to detect a driver's fatigue level. Svensson [29] even proposed an objective sleepiness scoring (OSS) method that relied on EEG. However, most of the driver biological measures are intrapersonal. The results of [71] showed that intrapersonal data had a good linear trend, while interpersonal data showed a different threshold. Bouchner [110] also showed that the EEG was very dynamic and very sensitive to outside factors. In addition, EEG patterns vary between individuals. Therefore, these two kinds of measures should be treated as rough ground truth indicators for other methods. Driver physical measures and driving performance measures are the most promising methods in the real driving context, because neither rely on intrusive measurements that might affect the driver.

For fatigue detection, the most popularly used parameter in driver physical measures is PERCLOS. However, one of the limitations of PERCLOS is that its prediction is good only when using large time intervals. Moreover, PERCLOS does not take into account the variability in human behavior, because blinking activity can significantly differ between individuals. Another challenge for driver physical measures is the robustness of the algorithm under all driving conditions (day and night, sunny and cloudy, etc.), because this type of method mainly relies on image processing. Many researchers have adopted infrared illumination techniques in image acquisition systems for three purposes. First, they minimize the impact of different ambient lighting conditions. Second, they allow the bright pupil effect to be produced, which makes eye detection easier. Third, because near-infrared is barely visible to the driver, it minimizes any interference with their driving. The "bright pupil" effect does benefit the eye extraction process, but it only works well under

some constrained lighting conditions. Moreover, in real driving scenarios, these constraints cannot be satisfied most of the time. In Bergasa et al. [50], three main illumination challenges were encountered: artificial light from elements outside the road (such as streetlights), vehicle lights, and sunlight, as shown in Fig. 8. The "bright pupil" effect will disappear under these conditions, which causes the eye detection to fail, and consequently influences the inattention detection. For example, sunlight and reflections from glasses could cause the performance to drop considerably, to 30% [50]. No matter how the hardware is adjusted, the "bright pupil" effect is not robust, especially in daytime [50] or when wearing glasses [27]. Even under constrained conditions, the reflection of the IR in the pupils varies by individual. Even with the same driver, the intensity depends on the gaze point, head position, and opening of the eye. Therefore, more reliable real-time eye detection algorithms are preferred over the "bright pupil" effect. As described in subsection "Driver Physical Signal Extraction", most studies have concentrated on image processing, and have estimated the driver physical parameters (e.g., gaze, face pose, mouth activity) quite roughly. Combining the image processing with some face mathematical models leads to more accurate estimation. Dong et al. [111] developed a real-time tracking kernel for stereo cameras to estimate face pose and face animation, including the movement of the eyelid, eyeball, eyebrow, and mouth, for driver inattention detection.

The advantage of driving performance measures is that the signals are meaningful and readily available. Moreover, the literature shows that they are useful for estimating driver fatigue, and show good promise in a real driving context. Although in subsection "Effects of Distraction", many researchers have found that driver distraction influences driving performance, few studies have utilized driving performance measures to detect distraction, with most concentrating instead on fatigue or abnormal detection.

One more issue that should be pointed out is that many of the researches claimed very high detection accuracies, which were true only for their particular hypothetical fatigue/distraction definitions. These hypothetical definitions usually covered a limited region of the whole fatigue/distraction definition. Without this condition, the accuracy rates had no meaning.

Because of the significant difficulties inherent in measuring driver attention, the magnitude, and particularly the safety implications, of driver distraction have been very difficult to determine. Indeed, unlike seat belt use, the driver's attention status cannot be categorized as "yes" or "no," but should be quantified in the same manner as blood alcohol level [24].

The factors that influence driver fatigue/drowsiness include greater daytime sleepiness, less sleep, a more difficult schedule, more hours of work, age, driving experience, cumulative sleep debt, the presence of a sleep disorder, and the time of day. This entry focused on real-time inattention detection technology, rather than on long-term sleep/wake regulation prediction technology. Biomathematical models that quantify the effects of the circadian and sleep/wake processes on the regulation of alertness and performance have been developed in an effort to predict the magnitude and timing of fatigue-related responses in transportation operations. These models of fatigue and

Driver Inattention Monitoring System for Intelligent Vehicles. Figure 8
Effects of external lights on the acquisition system. (**a**) Out-of-the-road lights effect. (**b**) Vehicle lights effect. (**c**) Sunlight effect. (**d**) Sunlight effect with filter [50]

performance typically use input information about sleep history, duration of wakefulness, work and rest patterns, and circadian phase to predict sleepiness, performance capability, and/or fatigue risk [47]. Researches on biomathematical models could enhance the confidence of the real-time estimation, because the former could be used to predict when the probability of sleepiness will become higher. For descriptions of these biomathematical models, see [47].

Systematical Design Consideration

This entry indicated that no single measure can be used to reliably detect inattentive driving. A combination of different measures is recommended, e.g., analyses of lateral control performance and eye blink patterns. According to the definition of driver distraction [2], when driver distraction occurs there should be some kind of distraction source that exists inside or outside of the vehicle. Therefore, fusing driver gaze information, vehiclears' ego state (steering, lane position, speed, and state of IVIS) with current road scenario (e.g., the type of road, weather conditions, and traffic density) will lead to a more comprehensive understanding and recognition of driver distraction.

The level of distraction associated with a given secondary task depends on the extent to which a driver is engaged in the task. Different secondary tasks have different requirements for concentration. Thereby, they have different levels of distraction. Based on the number of button presses and/or glances away from the forward road, Klauer et al. [112] defined three categories of secondary tasks: complex tasks, moderate tasks, and simple tasks. It was found that complex tasks carried three times the risk of involvement in a crash or near-crash as simple tasks or no secondary tasks. Specifically, for drivers performing complex secondary tasks, elevated likelihood ratios were found for the following conditions: dusk and unlighted darkness, rain, divided roads, and roads with grades (straight or curved). Therefore, identifying the environmental conditions is important to correctly assess the risk of the distraction.

Different environments contribute different risk levels for the same inattention state. Different environments could also induce the occurrences of different distractions. Specifically, Stutts et al. [3] reported that crashes associated with adjusting audio devices were more likely at night, those associated with moving objects inside the vehicle were more likely on nonlevel grades, and distractions involving communication with other occupants were more likely at intersections. Horne and Reyner [113] found various criteria associated with drowsiness-related accidents: the vehicle running off the road, no sign of braking, no mechanical defect, good weather, and the elimination of speeding. The NHTSA [27] reported some indirect cues: Accidents were more likely to be associated with the period from midnight to early morning, rural highways with a 55–65 mph speed limit, and fixed objects (trees, guardrail, highway sign, etc.). This, if these contextual cues could be taken into account when determining the risk level of an inattention occurrence and determining which countermeasure should be adopted, it would make the driver inattention monitoring system more reliable and acceptable.

Conclusion

This entry reviews the current state of the knowledge about driver inattention monitoring. Driver inattention increases driving risk, and has become a major factor in a considerable percentage of traffic accidents. Driver inattention has no universally accepted definition. However, based on a review of the literature, driver inattention is classified into two main categories, distraction and fatigue, each of which also contains a few subcategories. Summarily, distraction means the driver has the capability of paying attention, but their attention is shifted away from the primary driving task to some secondary task or attracted by some attractive object/event. Fatigue means the driver has exhausted his or her attention energy and cannot maintain sufficient attention for driving. The causes of distraction and fatigue are different, and they impose different influences on the driver and driving performance. Revealing these influences could help when selecting appropriate measures to develop a real-time inattention monitoring system. Recently, many commercial products relative to driver inattention monitoring have emerged. Auto companies such as Toyota, Nissan, Volvo, Mercedes-Benz, and Saab have installed driver inattention monitoring systems on their top-brand vehicles and/or are conducting researches on such

systems. A few third parties, like Seeing Machines and SmartEye, are providing camera-based nonintrusive tools for measuring driver physical signals such as gaze, head pose, and mouth activity. It should be pointed out that in most of the cases, neither the scientific and technological method behind, nor the exhaustive results of the performance can be provided for these commercial products.

Many articles have reported that these tools work well under constrained conditions, but are not robust under real driving conditions. Thus, there is still much progress to be made to improve the robustness and accuracy of the physical measuring tools. In the scientific literature, five different types of measures could be found for detecting driver inattention: (a) subjective report measures, (b) driver biological measures, (c) driver physical measures, (d) driving performance measures, and (e) hybrid measures. Although not suitable for a real life context, subjective report measures and driver biological measures could serve as some rough ground truth indicators. Because driver physical measures and driving performance measures have advantages and disadvantages, hybrid measures are believed to provide more reliable solutions, which will both accurately detect driver inattention and minimize the number of false alarms to promote acceptance of the system. After all, the goal of a driver inattention monitoring system is to reduce the driving risk. To obtain this goal, three distinct sources of data should be combined: driver physical variables, driving performance variables, and information from the IVIS. In addition to these variables, it is important to consider the characteristics of the driving environment (e.g., the type of road, weather conditions, and traffic density).

Bibliography

1. Ranney TA, Mazzae E, Garrott R, Goodman MJ (2000) Nhtsa driver distraction research: past, present, and future. Technical report, Transportation Research Center Inc
2. Lee JD, Young KL, Regon MA (2009) Driver distraction:. CRC Press/Taylor & Francis Group, London
3. Stutts JC, Reinfurt DW, Staplin L, Rodgman EA (2001) The role of driver distraction in traffic crashes. Technical report, American Automobile Association AAA Foundation for Traffic Safety
4. S.S.G. on Fatigue (2006) Canadian operational definition of driver fatigue. Technical report, STRID Sub-Group on Fatigue
5. Brill JC, Hancock PA, Gilson RD (2003) Driver fatigue: is something missing? In: Proceedings of the second international driving symposium on human factors in driver assessment, training and vehicle design, Park City, pp 138–142
6. T.R.S. for the Prevention of Accidents (2001) Driver fatigue and road accidents: a literature review and position paper. Technical report, The Royal Society for the Prevention of Accidents
7. Croo HD, Bandmann M, Mackay GM, Rumar K, Vollenhoven P (2001) The role of driver fatigue in commercial road transport crashes. Technical report, European Transport Safety Council
8. Horne J, Reyner L (1999) Vehicle accidents related to sleep: a review. Occup Environ Med 56(5):289–294
9. Nilson T, Nelson TM, Carlson D (1997) Accid Anal Prev 29(4):479
10. Endsley MR (1995) Hum Factors 37(1):32
11. http://www.chalmers.se/safer/driverdistraction-en/programme
12. Angell L, Auflick J, Austria A, Kochhar D, Tijerina L, Biever W, Diptiman T, Hogsett J, Kiger S (2006) Driver workload metrics project-task 2 final report. Technical report, U.S. Dept. of Transportation National Highway Traffic Safety Administration
13. Harbluk JL, Noy YI, Trbovich PL, Eizenman M (2007) Accid Anal Prev 39:372
14. Rantanen EM, Goldberg JH (1999) The effect of mental workload on the visual field size and shape. Ergonom 42(6): 816–834
15. Harbluk JL, Noy YI (2002) The impact of cognitive distraction on driver visual behaviour and vehicle control. Technical report, Ergonomics Division, Road Safety Directorate and Motor Vehicle Regulation Directorate
16. Hayhoe MM (2004) Infancy 6(2):267
17. May JG, Kennedy RS, Williams MC, Dunlap WP, Brannan JR (1990) Acta Psychol 75(1):75
18. Miyaji M, Kawanaka H, Oguri K (2009) Driver's cognitive distraction detection using physiological features by the adaboost. In: Proceedings of the 12th international IEEE conference on intelligent transportation systems, St. Louis, 2009
19. Liang Y, Lee JD (2010) Accid Anal Prev 42:881
20. Zhang H, Smith M, Dufour R (2008) A final report of safety vehicles using adaptive interface technology (phase ii: Task 7c): visual distraction. Technical report, Delphi Electronics and Safety
21. Itoh M (2009) Individual differences in effects of secondary cognitive activity during driving on temperature at the nose tip. In: Proceedings of the 2009 IEEE International conference on mechatronics and automation, Changchun, China, 2009
22. Wesley A, Shastri D, Pavlidis I (2010) A novel method to monitor driver's distractions. In: Proceedings of the 28th of the international conference extended abstracts on human factors in computing systems CHI EA 10 (2010), Atlanta, 2010
23. Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, Olmstead RE, Tremoulet PD, Craven PL (2007) Aviat Space Environ Med 78(5):B231
24. Ranney TA (2008) Driver distraction: a review of the current state-of-knowledge. Technical report, NHTSA
25. Zhou H, Itoh M, Inagaki T (2008) Influence of cognitively distracting activity on driver's eye movement during

preparation of changing lanes. In: SICE annual conference, Tokyo, 2008
26. Carsten OMJ, Brookhuis K (2005) Transp Res Part F: Traffic Psychol Behav 8:191
27. Eskandarian A, Sayed R, Delaigue P, Blum J, Mortazavi A (2007) Advanced driver fatigue research. Technical report, Report No. FMCSA-RRR-07-001. U.S. Department of Transportation Federal Motor Carrier Safety Administration
28. Dinges DF, Richard G (1998) Perclos: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Technical report, Federal Highway Administration, Office of Motor Carriers
29. Svensson U, (2004) Electrooculogram analysis and development of a system for defining stages of drowsiness. Master's thesis, Department of Biomedical Engineering, Linkoping University, Sweden
30. Hulbert S (1972) Effects of driver fatigue. In: Forbes TW (ed) Human factors in highway traffic safety research. Wiley-Interscience, New York
31. Mast T, Jones H, Heimstra N (1989) Effects of fatigue on performance in a driving device. Highway research record. Technical report, Driver Fatigue Research: Development of Methodology, Haworth, Vulcan, Triggs, and Fildes (eds) Accident Research Center, Monash University Australia
32. Kahneman D (1973) Attention and effort. Technical report, Prentice Hall, Englewood Cliffs
33. Yabuta K, Iizuka H, Yanagishima T, Kataoka Y, Seno T (1985) The development of drowsiness warning devices. In: Proceedings of the 10th International technical Conference on Experimental Safety Vehicles, Washington, DC
34. Dingus TA, Hardee L, Wierwille WW (1985) Development of impaired driver detection measures. Technical report, Department of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University, (Departmental Report 8504), Blacksburg
35. Elling M, Sherman P (1994) Evaluation of steering wheel measures for drowsy drivers. In 27th ISATA, Aachen, Germany
36. Zhong YJ, Du LP, Zhang K, Sun XH (2007) Wavelet Anal Pattern Recognit 4:1843
37. Skipper JH, Wierwille W, Hardee L (1984) An investigation of low level stimulus induced measures of driver drowsiness. Technical report, Virginia Polytechnic Institute and State University IEOR Department Report No.8402, Blacksburg
38. Stein AC (1995) Detecting fatigued drivers with vehicle simulators. Technical report, Drive r Impairment, Driver Fatigue and Driving Simulation, Hartley L (ed) Taylor & Francis, Bristol, pp 133–150
39. Safford R, Rockwell TH (1967) Highway Res Record 163:68
40. Riemersama JB, Sanders AF, Wildervack C, Gaillard AW (1977) Performance decrement during prolonged night driving. Technical report, Vigilance: Theory, Operational Performance and Physiological Correlates, Makie RR (ed) Plenum Press, New York
41. www.saabnet.com/tsn/press/071102.html
42. Nabo A (2009) Driver attention – dealing with drowsiness and distraction. Technical report, Saab Automobile AB
43. www.smarteye.se
44. www.testdriven.co.uk/lexus-ls-600h/
45. www.emercedesbenz.com
46. www.media.volvocars.com
47. Barr L, Popkin S, Howarth H (2009) An evaluation of emerging driver fatigue detection measures and technologies. Technical report, Volpe National Transportation Systems Center
48. www.seeingmachines.com
49. www.smivision.com
50. Bergasa L, Nuevo J, Sotelo M, Barea R, Lopez E (2006) IEEE Trans Intell Transp Syst 7(1):63
51. Liang Y, Reyes ML, Lee JD (2007) IEEE Trans Intell Transp Syst 8(2):340
52. Freund Y, Schapire RE (1997) J Comput Syst Sci 55:119
53. Lee J, Reyes M, Liang Y, Lee YC (2007) Safety vehicles using adaptive interface technology (task 5): algorithms to assess cognitive distraction. Technical report, The University of Iowa
54. Sathyanarayana A, Boyraz P, Hansen JHL (2008) Driver behavior analysis and route recognition by hidden markov models. In: Proceedings of the 2008 IEEE international conference on vehicular electronics and safety, Columbus, 2008
55. Kaida K, Takahashi M, Akerstedt T, Nakata A, Otsuka Y, Haratani T, Fukasawa K (2006) Clin Neurophysiol 117(7):1574
56. Ingre M, Akerstedt T, Peters B, Anund A, Kecklund G (2006) J Sleep Res 15(1):47
57. Craig A, Tran Y, Wijesuriya N, Boord P (2006) Biol Psychol 72:78
58. Jap BT, Lal S, Fischer P, Bekiaris E (2009) Expert SystAppl 36(2):2352
59. Chouvarda I, Papadelis C, Papadeli CK, Bamidis PD, Koufogiannis D, Bekiaris E, Maglaveras N (2007) Medinfo 12(2):1294
60. Yeo MVM, Li XP, Shen K, Smith EPW (2009) Safety Sci 47(1):115
61. Shen KQ, Li XP, Ong CJ, Shao SY, Smith EPW (2008) Clin Neurophysiol 119:1524
62. Lin CT, Wu RC, Liang SF, Chao WH, Chen YJ, Jung TP (2005) Circuits and systems I: regular papers, IEEE Trans 52(12):2726
63. Lin CT, Chen YC, Huang TY, Chiu TT, Ko LW, Liang SF, Hsieh HY, Hsu SH, Duann JR (2008) Biomed Eng IEEE Trans 55((5):1582
64. Liu J, Zhang C, Zheng C (2010) Biomed Signal Process Control 5:124
65. Skinner BT, Nguyen HT, Liu DK (2007) Classification of eeg signals using a genetic-based machine learning classifier. In: 29th annual international conference of the IEEE, Lyon. Engineering in Medicine and Biology Society, pp 3120–3123
66. Yang G, Lin Y, Bhattacharya P (2010) Inf Sci 180:1942
67. Chua CP, McDarby G, Heneghan C (2008) Physiol Meas 29(8):857

68. Damousis IG, Tzovaras D (2008) Intell Transp Syst IEEE Trans 9(3):491
69. Shuyan H, Gangtie Z (2009) Expert Syst Appl 36:7651
70. Balasubramanian V, Adalarasu K (2007) J Bodyw Mov Ther 11:151
71. Katsis CD, Ntouvas NE, Bafas CG, Fotiadis DI (2004) Assessment of muscle fatigue during driving using surface emg. In: Proceedings of the IASTED international conference on biomedical engineering, Article, Innsbruck, pp 259–262
72. Fan X, Sun Y, Yin B, Guo X (2010) Pattern Recognit Lett 31:234
73. Friedrichs F, Yang B (2010) Camera-based drowsiness reference for driver state classification under real driving conditions. In: 2010 IEEE intelligent vehicles symposium, San Diego, 2010
74. Pudil P, Ferrii FJ (1994) Floating search methods for feature selection with nonmonotonic criterion functions. In: Proceedings of the 12th international conference on pattern recognition, IAPR, Jerusalem, 1994
75. Sun XH, Xu L, Yang JY (2007) In: MIPPR 2007: automatic target recognition and image analysis; and multispectral image acquisition
76. Orazio TD, Leo M, Guaragnella C, Distante A (2007) Pattern Recognit 40(8):2341
77. Suzuki M, Yamamoto N, Yamamoto O, Nakano T, Yamamoto S (2006) Systems, Man Cybern 4:2891
78. Senaratne R, Hardy D, Vander B, Halgamuge S (2007) Lect Notes Comput Sci 4492:801
79. Rongben W, Lie G, Bingliang T, Lisheng J (2004) Monitoring mouth movement for driver fatigue or distraction with one camera. In: IEEE 7th conference on intelligent transportation systems, pp 314–319
80. Fan X, Yin BC, Sun YF (2007) Mach Learn Cybern 2:664
81. Vural E, Cetin M, Ercil A, Littlewort G, Bartlett M, Movellan J (2007) Drowsy driver detection through facial movement analysis. Springer, Berlin/Heidelberg
82. Kircher K, Ahlstrom C, Kircher A (2009) Comparison of two eye gaze based real-time driver distraction detection algorithms in a small-scale field operational test. In: Proceedings of the fifth international driving symposium on human factors in driver assessment, training and vehicle design, Big Sky, 2009
83. Pohl J, Birk W, Westervall L (2007) J Syst Control Eng 221(14)):541
84. Bergasa LM, Buenaposada JM, Nuevo J, Jimenez P, Baumela L (2008) Analysing driver's attention level using computer vision. In: 11th international IEEE conference on intelligent transportation systems, Beijing
85. Furugori S, Yoshizawa N, Iname C, Miura Y (2005) Rev Automot Eng 26(1):53
86. Farid M, Kopf M, Bubb H, Essaili A (2006) Safety Lit 1960:639
87. Takei Y, Furukawa Y (2005) IEEE international conference on systems, Man Cybern 2:1765
88. Wakita T, Ozawa K, Miyajima C, Igarashi K, Itou K, Takeda K, Itakura F (2006) IEICE Trans Inf Syst E89-D(3):1188
89. Torkkola K, Massey N, Wood C (2004) Driver inattention detection through intelligent analysis of readily available sensors. In: The 7th international IEEE conference on intelligent transportation systems, Washington, DC, 2004
90. Breiman L (2001) Random forests. Technical report, University of Califomia, Berkeley
91. Ersal T, Fuller HJA, Tsimhoni O, Stein JL, Fathy HK (2010) Model-based analysis and classification of driver distraction under secondary tasks. IEEE transactions on intelligent transportation systems 11(3):692–701
92. Sathyanarayana A, Nageswaren S, Ghasemzadeh H, Jafari R, Hansen JHL (2008) Body sensor networks for driver distraction identification. In: IEEE international conference on vehicular electronics and safety (ICVES), Columbus, 2008
93. Sathyanarayana A, Boyraz P, Hansen JHL (2011) Information fusion for robust 'context and driver aware' active vehicle safety systems. Inf Fusion 12(4):293–303
94. Doshi A, Trivedi M (2009) Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions. In: Intelligent vehicles symposium. IEEE, Xi'an
95. Weller G, Schlag B (2009) In: European conference on human centred design for intelligent transport systems
96. Markkula G, Kutila M (2005) Online detection of driver distraction – preliminary results from the aide project. In: Proceedings of the 2005 international truck and bus safety and security symposium, Virginia, pp 86–96
97. Tango F, Calefato C, Minin L, Canovi L (2009) Moving attention from the road: a new methodology for the driver distraction evaluation using machine learning approaches. In: HSI 2009, Catania
98. Fletcher L, Zelinsky A (2007) Driver state monitoring to mitigate distraction. Technical report. In: Faulks IJ, Regan M, Stevenson M, Brown J, Porter A, Irwin JD (eds) Distracted driving. Sydney, NSW: Australasian College of Road Safety pp 487–523
99. Tran C, Trivedi MM (2010) Towards a vision-based system exploring 3d driver posture dynamics for driver assistance: issues and possibilities. In: 2010 IEEE intelligent vehicles symposium, San Diego, 2010
100. Land MF, Lee DN (1994) Nature 369:742
101. Apostoloff N, Zelinsky A (2003) Int J Robot Res 5:28
102. Brandt T, Stemmer R, Rakotonirainy A (2004) IEEE Int Conf Syst Man Cybern 17:6451
103. Eren H, Celik U, Poyraz M (2007) Stereo vision and statistical based behavior prediction of driver. In: Proceedings of IEEE intelligent vehicles symposium, Istanbul, 2007
104. Su MC, Hsiung CY, Huang DY (2006) Syst Man Cybern 1:429
105. Ji Q, Yang XJ (2002) Real-Time Imaging 8:357
106. Cudalbu C, Anastasiu B, Radu R, Cruceanu R, Schmidt E, Barth E (2005) Signals Circuits Syst 1:219
107. Huang H, Zhou YS, Zhang F, Liu FC (2007) Wavelet Anal Pattern Recognit 3:1144
108. Zhu ZW, Fujimura K, Ji Q (2007) Data Sci J 6:636
109. Schmidt EA, Schrauf M, Simona M, Fritzsche M, Buchnerb A, Kincsesa WE (2009) Accid Anal Prev 41:1087

110. Bouchner P (2006) WSEAS Trans Syst 5(1):84
111. Dong Y, Hu Z, Uchimura K, Murayama N (2010) A robust and efficient face tracking kernel for driver inattention monitoring system. In: Intelligent vehicles symposium (IV), 2010 IEE
112. Klauer SG, Dingus TA, Neale VL, Sudweeks JD, Ramsey DJ (2006) The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical report, National Highway Traffic Safety Administration, Washington, DC
113. Horne JA, Reyner LA (1995) Br Med J 310(6979):565

Driving Under Reduced Visibility Conditions for Older Adults

Rui Ni
Department of Psychology, Wichita State University, Wichita, KS, USA

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Car Following in Fog
Collision Detection in Fog
Steering Control Under Reduced Contrast Conditions
Spatial Integration in Processing Optical Flow Information
Future Directions
Bibliography

Glossary

Contrast sensitivity The ability to distinguish different levels of luminance for a given display.

Definition of the Subject and Its Importance

An important and consistent finding regarding driving safety is that accident risk increases for older driver populations [1–3]. Evans [2] systematically examined data from the FARS (Fatality Analysis Reporting System) and found steady increases in accident fatalities and rate of severe crashes for older drivers who were older than 60. This increased accident rate occurred for both men and women and was independent of miles driven.

For many people, driving gives us a sense of independence. This becomes even more important for older population to maintain the quality of life. Unfortunately, natural age-related functional deteriorations cause decreased driving performance and increased crash risks. There are an increasing number of older adults on the road today compared to 10 years ago [4], and the older driver population will continue to increase in the next few decades. Thus, it is important to understand older adults' driving behaviors and their limitations so that we can ultimately help them to sustain a healthy and active lifestyle.

Owsley [5] identified four main functional impairments that contribute to decreased driving performance: (1) visual, (2) visual-cognitive, (3) cognitive, and (4) physical. These impairments can come at any age but are more common among the older population. Research on visual impairments and driving has mainly focused on the decline of visual acuity for older drivers and visual acuity assessment has been a standard screening test for the driver's license test. However, it has been shown that visual acuity is not a good predictor for accident risk [6]. Instead, dynamic acuity, contrast sensitivity, and useful field of view [7] were proved to be much better candidates to predict accident risks. Wood and Owens [6] found that real-world recognition performance of all age groups was independent of visual acuity, which however was seriously degraded during night driving, and this impairment was greater for the older participants. McGwin, Chapman, and Owsley [8] found that contrast sensitivity was significantly associated with the frequency of making left-hand turns, driving on high-traffic roads, driving during rush hours, driving alone, and parallel parking. Glass [9] showed that contrast sensitivity accounted for more age-related variability and was highly correlated with high sensory-demand tasks as compared with low sensory-demand tasks. Thus, it is critical to systematically examine older drivers' driving behaviors under different contrast conditions.

Introduction

The increased crash risk for older drivers could result from a number of age-related factors. These factors range from sensory processing and perceptual processing to attention and cognitive ability.

Age-related declines in sensory processing have been found in accommodation [10], contrast sensitivity [11–13], dark adaptation [14, 15], visual acuity [16, 17], spatial vision [18], and dynamic visual acuity [19]. Age-related changes in perceptual processing have been found in motion perception [20–24], optical flow [22, 25, 26] and depth perception [27, 28].

These types of age-related changes have important implications for driving safety. For example, declines in motion and depth perception can result in performance decrements in detecting impending collisions during decelerations [25] and of approaching objects [26]. Age-related declines in attention include performance decrements for both focused [29, 30] and divided attention tasks [31]. One issue that has been extensively studied is the decline in the useful field of view [7, 32–35]. Finally, age-related declines in cognitive ability include a consistent result in generalized slowing of cognitive processing [36, 37].

Safely traveling in a vehicle relies on proper skills and capabilities performing complex tasks which can be categorized into three types of activities – strategic, tactical, and control [38]. Strategic tasks involve the planning of the trip and the overall goals for the driving. Before we even begin to drive, we must plan the route to our new destination. Choosing a route can be influenced by many factors. For example, we may choose a particular route because it is the shortest route in distance. Other times, we may decide on a route which is the shortest in travel time, the most familiar one, or the most fuel efficient one. Once a route is determined, the driver attempts to follow that route, experiencing a phenomenon called wayfinding. Wayfinding is the ability for a person to navigate through and around an environment. Subtasks of wayfinding that are equally important to safe driving include maintaining adequate heading, vehicle velocity [39], and developing a cognitive representation of the environment [40].

Tactical tasks of driving focus on the decisions made for the on-hand goals of arriving at a destination. These tasks include speed selection, decisions to pass other vehicles, and lane selection. For example, detection of an imminent collision is crucial in safe driving. Imagine a driver waiting to make a left turn at an intersection. If there are oncoming vehicles approaching the intersection, the driver needs to decide whether or not it is safe to make the turn. In other words, the driver has to judge whether there will be a collision before the turn is finished, by estimating the time-to-contact (TTC) from a stationary standpoint. It will be safe to make the left turn if the TTC is greater than the time that it would take for the driver to finish the turn. Now imagine you are sitting in the upcoming vehicle approaching this intersection. You need to estimate the TTC of the left-turning vehicle from a moving standpoint. If a collision is imminent, you should make a quick judgment and execute the proper reactions in attempt to avoid the collision, by either swerving, braking, or even speeding up.

Control tasks focus on the operation of the vehicle. These tasks include maintaining safe distances between cars, maintaining speeds, maintaining lane positions, and steering the vehicle. For example, proper car following entails maintaining a safe distance between a driver's vehicle and the lead vehicle in front. A safe following distance provides an ample amount of time for the driver to react appropriately to sudden hazards. It is very often that drivers have to gather information about the surroundings with just one glance. DeLucia [41] found that young drivers were able to derive motion information in car following scenes fairly well, even when visual information was partially absent. However, older drivers extrapolated this information much slower than their younger counterparts.

Lane keeping and lane changing are similar driving tasks. Lane keeping is the act of keeping the vehicle within the lane boundaries as the road and environment change, e.g., on a curved road. Lane changing is also necessary to avoid certain situations (e.g., avoid colliding into a car merging into the freeway) as well as maintaining safe, efficient wayfinding. Recent study by Macuga and colleagues [42] investigated whether people could perform a proper lane change with little or no visual information. When the needed heading direction was made implicit, participants were able to change lanes. Further, drivers were able to navigate properly even when all visual information was removed, by using inertial cues in a portable virtual reality vehicle. Again, research has demonstrated that drivers are able to perform certain driving tasks, such as lane changing, with little or no visual information within a certain amount of time.

Driving performance can also be assessed in terms of steering control error. Steering control is a basic task in driving that is to control the heading direction of the automobile. Steering control is also essential for many driving tasks that are previously discussed, e.g., car following, lane changing and keeping, and collision avoidance. Perceiving correct heading direction is necessary to maintain or adjust locomotion and may be the lowest-level steering control task. Therefore, it has been the primary driving performance measure for a large body of driving-related literature.

In the current article, a series of driving studies are reviewed that investigated age-related decrements in driving performance, especially in low visibility condition (i.e., in low contrast driving scenes). Several essential driving tasks are examined, including car following, collision detection, and steering control. In addition, the age-related spatial integration mechanism is proposed to account for age-related deficiency in driving in optical flow field. In the end, future research directions are considered briefly on both the advances of new technology and improving driving performance through training.

Car following in Fog

In this part age-related differences are examined in sensory and perceptual processing for a task important for driving safety – car following. Effective car following allows drivers to maintain safe distances that are necessary to take proper executions in case of emergency. Failure to correctly respond to lead vehicle (LV) speed changes can have serious consequences for the safety of the driver. For example, if a driver in a following vehicle fails to respond to a reduction in LV speed then the headway distance between the following and lead vehicle is reduced. This can result in a following distance that is too close (i.e., does not allow for sufficient response time should the LV suddenly decelerate), leading to an increased risk of a crash.

Previous research on car following [43, 44] has assumed that drivers have precise information regarding headway distance and speed of the lead and following vehicle. A limitation of this research is that drivers do not have access to this precise information. Instead, drivers can estimate distance and speed based on available sources of visual information. Recently, Andersen and Sauer [45] proposed a new model for car following, referred to as the DVA (driving by visual angle) model, based on the visual angle and speed information available to a driver. The DVA model consists of two components – one component provides information for distance perception (based on visual angle of the LV) and the second component provides information useful for speed perception (based on instantaneous changes in LV visual angle). The results of their study indicate that the DVA model, as compared to other car following models based on precise headway distance and LV speed, could better predict driver performance in both simulator and real-world driving conditions. In a related study, Andersen and Sauer [46] showed the use of the driving scene information in car following. The results indicated greater accuracy in driving performance when the surrounding scene was visible as compared to conditions when it was not visible. They argued that the surrounding scene is useful for specifying edge rate information, which is used to estimate the speed of the driver's vehicle.

Recent study by Ni, Kang, and Andersen [47] examined two hypotheses concerning age-related changes in car following performance. Previous studies have shown age-related differences in the use of scene information to judge distance and layout of a scene [48]. This finding suggests that older drivers, as compared to younger drivers, will have decrements in perceived distance to the LV. This hypothesis is referred to as the aging and distance perception hypothesis. Previous studies have also found age-related declines in judging speed of the driver's vehicle [25] and of the approaching objects [32, 49]. These findings suggest that older drivers, as compared to younger drivers, will have decrements in perceiving self-speed and relative speed change between the lead and following vehicle. This hypothesis is referred to as the aging and speed perception hypothesis.

In addition to examining age-related differences in car following performance, Ni's study [47] examined a set of environmental conditions, the presence of fog, that are likely to be problematic for older drivers. Fog reduces the overall contrast and visibility of the driving scene, with the magnitude of reduced visibility increasing as a function of distance. As a result, the ability to see detail of the driving scene is reduced as a function of

the distance between the driver and objects in the scene. Epidemiological studies of older driver crash rates have found increased crash risk for older drivers under reduced visibility conditions due to weather or dusk/nighttime conditions [1, 50–52]. A more recent study [53] found reduced car following performance for college-age drives under simulated fog conditions.

As discussed earlier, it is well documented in the literature that contrast sensitivity is reduced with increased age [11–13]. Indeed, studies have found that for photopic vision (daylight conditions) there is a 41% reduction of contrast sensitivity for 70 year old subjects as compared to 20 year old subjects in detecting mid- to high level spatial frequency targets [13]. These results suggest that older drivers are likely to have poorer car following performance than younger drivers because of reduced visibility of the LV under dense fog conditions.

Previous research [53, 54] has found decreased car following performance (failure to maintain following distance) under simulated fog conditions. Two factors may impact car following performance under foggy conditions. First, the reduced visibility of the scene may result in a compression of the overall perceived depth of the driving scene. This situation is likely to occur as the reduction in contrast of the surrounding scene will remove information important for perceiving scene depth such as texture gradients and linear perspective [55]. If the perceived depth is compressed then smaller headway distance at high fog density conditions is expected. Second, reduced visibility of the scene may result in increased difficulty in estimating speed. Previous research has found that edge rate information (the rate at which local edges cross a fixed reference point in the visual field) is important for determining the perceived speed of vehicle motion [56] and is used by drivers for tasks such as braking [57, 58]. If the reduced visibility of the scene from fog decreases the visibility of edge rate information in the driving scene then greater error in tracking changes in LV speed is expected.

The rest of this section then focuses on the study by Ni and colleagues [47]. In this study, Eight college students (age mean and standard deviation of 21.0 and 2.6, respectively) and eight older subjects (age mean and standard deviation of 72.6 and 4.6, respectively) participated in the study and all reported normal or corrected-to-normal vision and were currently licensed drivers. All drivers had experience driving in fog and reported driving at least 3 days per week.

The study was conducted in a fix-based driving simulator, in which the roadway in an urban setting consisted of three traffic lanes (representing a three lane one way road) with the driver and LV located in the center lane. Drivers were presented with a car following scenario in which the LV varied its speed according to a sum of three equal-energy sinusoids (i.e., the peak accelerations and decelerations of each sine wave in the signal were equivalent). At the beginning of each trial run, drivers were given 5 s of driving at a constant speed 18 m behind the constant speed LV. Drivers were instructed that the headway distance during this phase of the trial was the desired headway distance. Following 5 s a tone sounded to indicate to the driver that the LV speed would vary, which was based on the sum of three non-harmonic sine-wave frequencies. This signal does not repeat and thus prevents the driver from anticipating changes in LV speed.

To simulate realistic effects of fog, specific fog density values were selected to represent a range of conditions from high visibility (0.0 fog condition) to low visibility (0.16 fog condition). The low visibility condition was selected based on informal observations indicating that under this condition the visibility of the LV was considerably reduced at the desired following distance of 18 m. The LV was visible to all drivers under the simulated fog conditions at the predetermined driving distance.

Car following performance was assessed using a variety of measures that examine overall performance changes based on distance information and speed information. Specifically, mean and variance of distance headway are based on distance perception whereas RMS speed error is based on speed perception. If the age and distance perception hypothesis is correct, then age-related declines in measures based on distance perception are expected. If the age and speed perception hypothesis is correct, then age-related declines in measures based on speed perception are expected. In addition, if older drivers have greater difficulty in detecting and responding to speed changes then age-related declines should be greater as the overall speed of the LV is increased.

An important issue in car following is how to quantify a safe following distance. Time headway (THW) is a measure that is derived by the ratio of headway distance to the velocity of the driver vehicle. This measure indicates the time between two vehicles passing the same point traveling in the same direction and thus has been used as an indication of a safe margin between the driver and lead vehicle. In Ni's study they required drivers to maintain a fixed distance of 18 m across the three speed conditions. Thus, it is natural that THW would vary as a function of speed. The use of THW allows one to determine changes in safe driving performance as a function of fog density. For example, consider car following at a specific constant speed. If drivers have difficulty in maintaining a safe margin as a function of fog then increased fog density should result in a decrease in THW. If older drivers have greater difficulty than younger drivers under foggy conditions in determining distance and speed then they may follow at a closer distance resulting in a decrease in THW – an indication of increased crash risk.

The results showed that the mean distance headway (distance between driver and lead vehicle in meters) for the 0.0, 0.04, 0.08, 0.12, and 0.16 simulated fog density levels were 19.3, 20.1, 19.3, 18.2, and 17.9 m, respectively. Mean distance headway for the 40, 60, and 80 km/h speeds were 17.7, 19.2, and 20.0 m, respectively. The result also showed that older drivers maintained a slightly greater headway distance for the no fog (0.0 fog density) condition. However, at higher fog density levels and at increased speed older drivers maintained a closer headway distance than younger drivers. The largest age-related difference in distance headway occurred at the intermediate speed (60 km/h) and highest fog density condition. However, a notable exception was the highest fog density condition and highest speed. For this condition older drivers had a greater following distance than younger drivers. This result is likely due to a change in strategy employed by older drivers. Specifically, older drivers increased following distance at high speeds to minimize the likelihood of a collision because of difficulty in perceiving changes in the visual angle of the LV due to reduced contrast. In addition, the result indicated that older drivers had greater headway variance (mean variance of 25.6) compared to younger drivers (mean variance of 10.1). Analysis of RMS speed error showed greater RMS error for older drivers (mean RMS error of 6.17 km/h) as compared to younger drivers (mean RMS error of 4.83 km/h).

To examine the effects of fog on the safety margin for car following time headway (THW) data was derived. THW varied as a function of the average speed of the LV with THW decreasing at increased average speed of the LV. The result showed decreased THW as a function of fog density. These results showed that at the intermediate speed (60 km/h) THW decreased for both older and younger drivers with an increase in fog density. However, older drivers had the largest decrease in THW at the highest fog condition. This result indicates that older drivers, as compared to younger drivers, had a significant reduction in the safety margin at the highest fog density level. At the highest speed (80 km/h) both older and younger drivers had a decrease in THW with an increase in fog density with the exception of the highest fog density condition. For younger drivers THW continued to decrease at the highest fog density condition. However, for older drivers THW increased at the highest fog density condition. This result suggests that older drivers increased the following distance to increase the safety margin.

As discussed earlier, two hypotheses were examined in this study concerning age-related differences in car following performance under foggy conditions. According to the aging and distance perception hypothesis, older drivers, as compared to younger drivers, will have decrements in perceived distance to the LV. According to the aging and speed perception hypothesis, older drivers, as compared to younger drivers, will have decrements in perceived speed and relative speed change between the lead and following vehicle. The results of this study provided evidence in support of both of these hypotheses. With regard to distance information, they found that older drivers, as compared to younger drivers, followed at a closer headway distance with an increase in fog density. In addition, although mean headway distance varied for both younger and older drivers as a function of speed, older drivers showed a greater change as a result of speed. This effect was especially pronounced for the highest fog density conditions examined. They also found greater variance in distance headway for older drivers,

as compared to younger drivers. These results provide evidence in support of the age and distance perception hypothesis.

With regard to speed perception the results indicated that older drivers had greater RMS speed error than younger drivers. In addition, older drivers had lower squared coherence scores than younger drivers, indicating that older drivers had poorer performance in tracking local variations in speed at specific frequencies of LV speed. These results provide evidence in support of the age and speed perception hypothesis. The results, considered together, suggest that older drivers have decreased car following performance as a result of difficulty in judging both speed and distance.

An important finding in this study was the interaction of age, speed, and fog density for the distance headway measure. Of particular interest were the age-related performance differences for the 60 km/h speed condition under the highest fog density condition. Under this condition, older drivers followed at a very close distance and shorter THW. The authors believed that older drivers followed at a close distance because reduced visibility due to fog increased the difficulty to perceive LV speed changes. Collision risk can be assessed by deriving the THW. The shorter the THW, the less is the time available to the driver to avoid a collision. For the 60 km/h/0.16 fog density condition younger drivers had an average time gap of 1.1 s. However, older drivers had a time gap of 0.87 s, which represents a 21% reduction in available response time to avoid a collision. Given the well-documented finding of slower reaction time with age, this result suggests that older drivers may be at considerable risk of a collision under high fog density conditions at moderate speeds.

Collision Detection in Fog

In addition to car following, there is another important perceptual task for driving safety – collision detection. Failure to correctly detect an impending collision event can greatly increase accident risk for drivers. For example, if a driver in a vehicle fails to see an upcoming vehicle early enough or does not see it at all when making a left turn, then a crash might occur resulting in a severe injury or even death for the driver. Accurate detection of an impending collision is the basis of making appropriate responses, such as steering or braking to avoid a collision. DeLucia et al. [59] studied the age-related differences in detecting collision events when an object was moving toward the observer at a constant speed. They found an age-related decrement in sensitivity for older females but not older males. Andersen and Enriquez [26] found that older observers, as compared to younger observers, were less sensitive to collision events and required more time to process the visual information.

As reviewed earlier, previous studies of older driver have shown increased crash rates under reduced visibility conditions due to weather or dusk/nighttime conditions. Specifically, the presence of fog reduces the overall contrast and visibility of the driving scene, which results in reduced visible details as a function of increasing distance. Such reduced visibility of driving scene may lead to higher injury and death rates in accidents [60]. Driving simulation studies have found evidence of decreased driving performance under simulated fog conditions. Recent study [47] reviewed above on car following task has shown reduced performance under simulated fog conditions for older drivers compared with younger drivers. As discussed earlier, contrast sensitivity is reduced with increased age. These findings suggest that older observers may have more difficulty than younger observers in recovering visual information needed to correctly detect impending collision events under dense fog conditions.

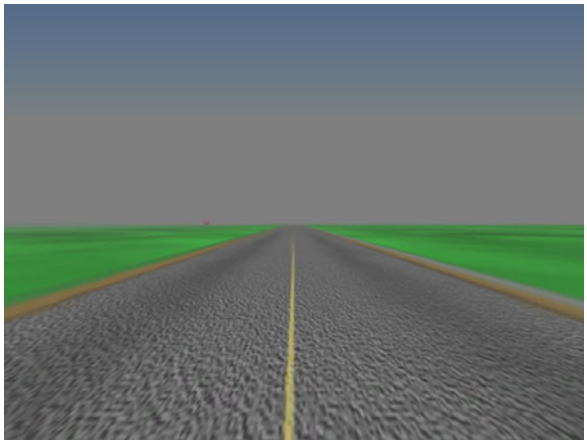
In this section, I will review a most recent research [61] which investigated detecting collision events when an object was approaching the observer at a constant speed on a linear trajectory under foggy conditions. Two factors might affect collision detection performance in fog. First, the reduced visibility of the scene may result in a reduced visibility of bearing angle of the approaching object, which is crucial to judge the collision event on linear trajectory [62]. This situation is likely to occur as the reduction in contrast of the surrounding scene will remove information important for perceiving object distance such as the horizon, texture gradients, and linear perspective. If the information on object position in the visual field is compromised, then reduced collision detection performance at dense fog condition is expected. Second, reduced visibility of the scene may result in increased difficulty in estimating speed. Previous research has

found that edge rate information (the rate at which local edges cross a fixed reference point in the visual field) is important for determining the perceived speed of vehicle motion [56] and is used by drivers for tasks such as braking [25, 58]. When moving in the driving scene, drivers need to predict not only the motion path of the approaching object but also the vehicle position after a certain amount of time, which is dependent of estimated speed. Thus, it is expected that the decreased visibility of edges in the driving scene under dense fog condition will decrease collision detection performance.

In this study, they investigated the effects of fog on collision detection performance when optical variables (i.e., visual angle and change in visual angle of the object, and the relative speed between the object and the vehicle) were constant. Drivers were presented with a driving simulation scene of a straight roadway in a suburban setting. A spherical object traveled on a linear trajectory toward the vehicle for 9 s before it either collided with or passed the vehicle. The vehicle was either stationary in Experiment one or moved straight ahead at a constant speed in Experiment two while the relative speed between the object and the vehicle remained constant. Only a part of the travel path was presented and drivers were instructed to make a judgment whether the object would collide with the vehicle or not by pressing one of the two keys on the keyboard.

In the first experiment, three independent variables were examined: Age (younger and older drivers), simulated fog density (no fog, mild fog, medium fog, or dense fog, corresponding to 0, 0.08, 0.16, or 0.24 in density, respectively), and the time-to-contact (TTC) on the last frame of the display (2, 4, or 6 s).

In the second experiment, the stimuli were the same as in first experiment except that the vehicle moved at one of the three speeds (20, 50, and 80 km/h). Since the relative speed between object and observer was kept constant, which is 90 km/h, the object moved at a speed of 70, 40, or 10 km/h, respectively, depending on the vehicle moving speed. Four independent variables were examined: Age (younger and older drivers), simulated fog density (no fog or dense fog, corresponding to 0 or 0.24 in fog density, respectively), moving speed of the vehicle (20, 50, or 80 km/h), and TTC (2, 4, or 6 s).



Driving Under Reduced Visibility Conditions for Older Adults. Figure 1
Example of the display in dense fog condition. A bright red spherical object moves along a linear trajectory toward the observer, which might or might not result in colliding with the observer

The displays simulated a 3D scene consisting of a roadway, roadway strip, textured ground, and an approaching bright red spherical object (See Fig. 1). The object motion trajectories were generated using the same method employed by Ni and Andersen [62]. The initial position of the object was chosen randomly from an arc (+/− 20° from the center of the display) at a fixed distance of 223 m from the viewpoint. The end position of the object was 50, 100, or 150 m from the view point for TTC of 2, 4, and 6 s, respectively. The 3D velocity for collision and non-collision events was constant. The simulated foggy effect was achieved using the same method described above. The low visibility condition was selected based on informal observations indicating that under this condition the visibility of the object was considerably reduced while still visible at the original position. All observers reported during debriefing that they were able to see the sphere in all conditions.

The two experiments were conducted in a darkened room. Observers viewed the monitor binocularly through a collimating lens with their head in a chin rest. They were informed that they would be shown a series of displays consisting of a 3D scene with a single object in a distance that was moving toward the

observer. Their task was to determine whether the moving object would hit the observer. Observers were next shown a series of demonstration displays of collision and non-collision stimuli in which the complete motion trajectory was displayed. After observers understood the task, they were presented with eight practice trials with complete 9-second events, half of which simulated a collision, and were asked to indicate whether or not the object was on a collision trajectory by pressing one of the two designated keys on a standard keyboard. Subjects pressed the number "6" key for non-collision responses and the number "4" key for collision responses. No feedback was given during the practice trials and experimental trials. Subjects were required to correctly identify seven out of the eight practice trials before proceeding to the experimental trials. Following the practice trials subjects were instructed that the experiment would include brief presentations of the collision and non-collision events. If they were uncertain of whether the display was a collision or non-collision event, they should make the best judgment possible.

The average proportion of hits (collision response for trials that simulated a collision) and false alarms (collision response for trials that did not simulate a collision) was calculated for each subject in each condition and used to derive sensitivity (d') and response bias (β) measures[63]. In conditions where observers performed perfectly (100% hit rate or 0% false alarm rate), d' and β were calculated based on the assumption that one error was made.

The analysis of d' values indicates that younger observers (mean $d' = 3.43$) had greater sensitivity than older observers (mean $d' = 2.78$). The increase of fog density resulted in decreased sensitivity for older observers but not for younger observers. When the TTC was 2 s, older observers performed as well as younger observers. As the TTC increased, the sensitivity for both age groups decreased. However, older observers' sensitivity decreased much faster than that of younger observers.

The analysis of β values suggests that both age groups were equally likely to make a collision judgment and the fog density did not bias the observers' likelihood to report a collision event. It was also found that observers were more likely to make a collision judgment with increased TTC.

The results of the second experiment are consistent with those in the first experiment, indicating that younger observers (mean d' = 2.65) had greater sensitivity than older observers (mean d' = 2.13). It was shown that with increased TTC the sensitivity for both age groups decreased. However, older observers' sensitivity decreased faster than that of younger observers. The analysis of β values suggests consistent results as in the first experiment that when fog was simulated or when the TTC increased, observers were more likely to respond that a collision would occur.

According to the contrast and bearing angle hypothesis, older drivers, as compared to younger drivers, will have more difficulty in perceiving constant bearing information under low contrast condition caused by fog. The results in Experiment 1 provide evidence supporting this hypothesis. It was found that older drivers, as compared to younger drivers, had significant decrements in sensitivity to detect an imminent collision event in dense foggy condition. In addition, although a decreased performance was found for both younger and older drivers with decreased display duration, older drivers showed greater decrements as a result of shorter display duration. This is a finding consistent with previous research (Andersen & Enriquez, 26). This result is probably due to the increased difficulty in discriminating bearing angle in driving scene with reduced contrast in distance.

Since reduced contrast in driving scene in dense fog condition caused decreased sensitivity of detecting collision events, one may ask whether this decrease can be predicted by contrast sensitivity. To address this concern, a linear regression between contrast sensitivity and the sensitivity of collision detection in fog was conducted for both age groups. The results indicated that contrast sensitivity was not a significant predictor for sensitivity of collision detection for either younger observers ($r^2 = 0.09$, $p = 0.37$) or older observers ($r^2 = 0.06$, $p = 0.46$). There are two explanations which might account for this result. First, the contrast sensitivity for each observer was measured using Pelli Robson test [64], which may not have sufficient precision to assess individual differences. For example, 10 out of 11 younger observers in this study had the same contrast sensitivity whereas their d' ranged from 2.38 to 3.24. An important issue for future research would be to precisely derive contrast sensitivity

thresholds. Second, the Pelli Robson test only measured contrast sensitivity in static scenes whereas our stimuli were in motion. Observers might have different contrast sensitivity when viewing motion stimuli as compared to when viewing static stimuli. A similar result has been found for studies examining sensitivity to binocular disparity [65]. Specifically this research has found that observers who were stereo-blind using static stimuli had sensitivity to disparity when the stimuli were moving. If similar conditions exist for low contrast stimuli, then the assessment of contrast sensitivity with static stimuli would not be predictive of sensitivity to low contrast moving stimuli.

With regard to edge rate information, it was found that both younger and older drivers had considerable decrease in their sensitivity to collision events when the vehicle is moving. The results in Experiment 2 were consistent with those in Experiment 1 in that decreased performance was found with decreased display duration (i.e., greater TTC). Considered together, these results suggest that under inclement weather conditions, such as fog, older drivers may be subject to an increased crash risk due to a decreased ability to detect impending collision events. One important finding in this study was the interaction of age and display duration, specifically the age-related difference for the 3 s condition. Under this condition, older drivers, as compared to younger drivers, had significant decrements in detecting the collision event which suggests a serious collision risk for older drivers. The longer it takes a driver to detect imminent collision events, the less time the driver has to take appropriate response, e.g., to decelerate or steer the vehicle. Given the well-documented finding of slower reaction time and reduced divided attention capacity with age this result suggests that older drivers may be at considerable risk of a collision under high fog density conditions.

Steering Control Under Reduced Contrast Conditions

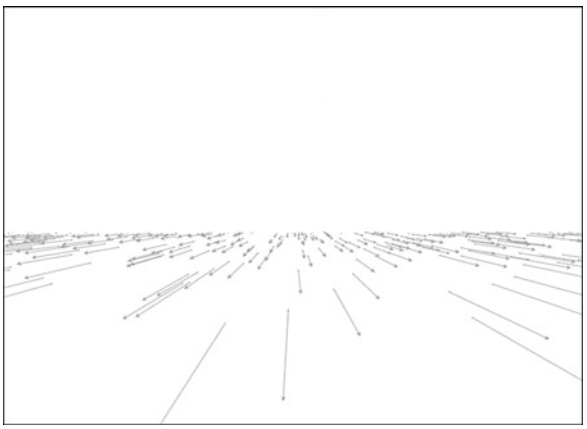
Besides accelerating and decelerating the vehicle, the other control task that a driver performs during driving is steering control, which is to steer the vehicle to move in desired direction. To successfully steer the vehicle, one needs to first perceive the current status and moving direction from the visual information in the driving scene. Failure to accurately detect changes in the path of motion and make corrections to the vehicle's path of motion could have serious consequences for driver safety. One source of information that has been extensively studied for the perception and control of steering is optical flow – the perspective transformation of the optic array [66, 67]. Research on optical flow has demonstrated the usefulness of this information for the perception of heading [68, 69], the perception of self-motion [70], and the perception of egospeed [56]. A critical assumption of optical flow research is that the spatial and temporal characteristics of the stimuli result in a perception of apparent motion.

Previous research has demonstrated age-related decrements in the perception of motion and in the use of optical flow information for the perception of 3D shape [22]. In this part, I will introduce a most recent study on age-related differences in the use of optical flow for steering control under low contrast conditions. In the study by Nguyen et al. [71], drivers were presented with computer generated displays simulating forward vehicle motion through a 3D scene of random dots on a ground plane. The horizontal position of the vehicle was perturbed according to a sum of sinusoidal functions, and drivers were asked to steer the vehicle to maintain the initial straight path of motion.

Previous research has shown that the use of optical flow information for perceiving the path of observer motion requires the spatial integration of velocity information. Recently, it was found that older observers had a decreased ability to spatially integrate velocity information to perceive motion defined edge boundaries ([72], will be reviewed in details below). Nguyen et al. [71] proposed that older drivers, as compared to younger drivers, would show a decreased sensitivity to spatially integrate velocity information for steering control from optical flow, especially in low visibility conditions. This was referred to as the spatial integration hypothesis. To test this hypothesis they manipulated the dot density display (resulting in a greater spatial separation of the velocity information) and contrast of the dots display. If the spatial integration hypothesis is correct, then age-related differences in steering performance should be more pronounced at low as compared to high dot densities and at low-contrast conditions as compared to high-contrast conditions.

In their study, Nguyen et al. manipulated the number of dots in the optic flow field (25 or 325), and contrast level (low, medium, and high level, representing 0.15, 0.2, and 0.25 in Michelson contrast, respectively). The displays simulated driving through a 3D array of dots located on a ground plane, as illustrated in Fig. 2. The simulated speed was 72 km/h. The horizontal position of the dots pattern was perturbed according to a sum of three prime sine-wave functions. The frequencies, amplitudes, and phases of the sinusoidal functions were chosen in a similar way as in Ni et al [47] so that the output of the sum of three sine-wave functions was always zero at the beginning of each trial and unpredictable. The scene consisted either 25 or 325 randomly positioned white dots on a black background to form the ground plane. The background had an average luminance of 16 cd/m^2. The dots were randomly positioned at the beginning of each trial, and then moved to the observer. Any dot that moved out of view port would be randomly repositioned at the far end of the ground plane.

Drivers were instructed that their task was to maintain a fixed heading direction relative to the moving dots as if driving down a straight roadway with lateral wind gusts perturbing their positions on the roadway. Steering control performance was assessed by calculating RMS (root means square) steering error for each driver on each trial in each condition. The average results for RMS error in their study revealed that although both older and younger drivers had less steering error with a greater number of dots in the flow field, younger drivers had significantly less steering error (9.9) than older drivers (12.4). Under low dot density condition, in which less optical flow information was presented, both younger and older drivers showed decreased steering performance with decreased contrast of the scene. However, older drivers' performance deceased much faster with decreased contrast, as compared with younger drivers. This result suggests that older drivers are more prone to accident risk than younger drivers under low contrast condition, especially with less optical flow information.



Driving Under Reduced Visibility Conditions for Older Adults. Figure 2
Low contrast optic flow showing straight ahead moving direction. In the actual experiment, the contrast was reversed so that *white dots* were displayed on a black background

Previous study by Ni and Andersen [73] found older drivers, as compared to younger drivers, relied more on optical flow than landmark information for their steering control task. One possible explanation for such result is that in their displays, only one color coded dot was present at a time which provided landmark information. In order to use landmarks for steering control, a driver must encode the spatial location of landmark positions and use this positional information as reference information to maintain the initial path of motion. Older drivers may have difficulty encoding and using landmark position information, which results in reduced accuracy in steering control, especially when small amount of landmark information is present. If this hypothesis is true, with increased landmark information there should be an increase in the steering performance for older drivers. However, older drivers might show less improved driving performance, which results from the extra landmark information added to optical flow field, as compared to younger drivers.

In a related study by Ni et al. [74], they used similar experimental setting as in the study by Nguyen et al. [71], except that different numbers of landmark dots were presented in the optic flow field (0, 1, 4, 7, and 10). Their result showed that younger drivers had significantly less steering error (5.8) than older drivers (9.9). Both older and younger drivers had less steering error with a greater number of landmark dots in the flow field. However, younger drivers, as compared to older drivers, were more efficient at using landmark information to improve steering control.

Overall, these findings indicate that older drivers may be more reliant on optical flow information for controlling a vehicle and may have a reduced ability to use alternative sources of information for steering control. The decreased reliance of landmark information for older drivers, when optical flow information is reduced, may result from age-related deficits in attention. Previous research [75] found that older subjects had greater difficulty than younger subjects in ignoring distracting items when scanning a display, presumably a result of difficulty in disengaging attention to irrelevant information. In Ni et al's study [74], older drivers may, under reduced optical flow conditions, focus their attention on the landmarks and have difficulty disengaging their attention from the landmarks to scan other parts of the scene. However, when more landmark information was presented on a dense optical flow ground plane, less steering control error occurred for older drivers since they can rely more on optical flow information in this condition.

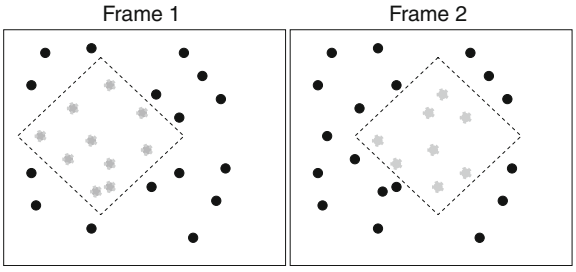
Spatial Integration in Processing Optical Flow Information

Visual processing of optical flow information requires the integration of information over space and time. This section attempts to explain age-related difference in using optical flow information in terms of spatial integration. Now consider how the visual system perceives a 2D form from fragmented image information. In order to perceive the form, the visual system must combine local fragments (e.g., line segments) across space to recover the shape of the form. Information for perception can also be available from different time intervals. For example, consider the perception of a moving object partly obscured by nearer objects. In order to perceive the form, the visual system must integrate the different parts of the object visible at different time intervals. These examples highlight the role of spatial and temporal integration in vision.

Previous studies of neuronal activity in the visual system have shown evidence of spatial and temporal integration. For example, retina cells respond to information present over a limited region of the visual field [76]. The visual system combines or spatially integrates information from local receptive field regions via intracortical connections from other cells within a cortical region [77–79] and from cells in higher cortical regions [80–82]. Spatial integration has been demonstrated to be important for visual processing including motion perception [83] and form perception [84–86].

Neurophysiological studies have also examined the role of temporal integration by the visual system. For example, studies have examined the role of temporal integration for motion perception in primary visual cortex (V1) and extrastriate motion area (MT/V5) [87]. It has been argued that temporal integration at a neural level is due to the synchronous activity of groups of neurons rather than the firing rate of a single neuron [88]. Specifically, temporal integration has been shown to be important in visual processing in motion perception [89] and form perception [90–93].

Recently, Andersen and Ni [72] used a novel perceptual task to examine whether age-related declines in visual processing resulted from changes in spatial or temporal integration. To assess this issue, they used a visual task that requires both spatial and temporal integration – the recovery of 2D shape from kinetic occlusion. Kinetic occlusion can specify 2D shape by providing information regarding edge boundaries of an object based on the accretion and deletion of texture over time. Consider an opaque object positioned in front of a background (see Fig. 3). If the object and background have identical random texture (randomly positioned dots) then the object boundaries cannot be seen. However, when the object starts moving (or the background starts moving) the accretion and deletion of the background texture due to occlusion can be used



Driving Under Reduced Visibility Conditions for Older Adults. Figure 3
Illustration of 2D form perception results from occlusion when the form moves from frame 1 to frame 2. Note the *stars* in low contrast represent the *dots* that are occluded by 2D form above (a *diamond* in this illustration)

to reveal the boundaries of the object. In order to recover the object boundary, the visual system must integrate the accretion and deletion of texture over space (discrete samples of the edge boundary are specified by the local disappearance/reappearance of texture) and time (the accretion and deletion of texture occur in local intervals of time).

Previous research [90] examined the importance of kinetic occlusion for specifying the shape of 2D objects. College-age subjects were presented with displays of a moving opaque object with random dot texture against a random dot texture background. The velocity and texture of the display were varied and subjects were asked to identify which of four shapes were present. The results indicated that an increase in velocity and density resulted in greater accuracy in shape identification. The results also indicated that density was the primary factor in determining performance. An important issue examined in their study was whether performance was based on the rate of occlusion events (the number of occlusion events per unit time). To examine this issue, they compared performance across conditions in which velocity and density changed proportionally resulting in a constant rate of occlusion events. For example, an occlusion stimulus with a velocity of 0.8 deg/s and a density of 1.2 dots/deg^2 was compared to an occlusion stimulus with a velocity of 1.6 deg/s and a density of 0.6 dots/deg^2. These two conditions have identical rates of occlusion events (i.e., disappearance/reappearance of texture). The results indicated that performance varied across conditions with a constant rate of occlusion events, with performance increasing due to an increase in density. These findings led Andersen and Cortese to conclude that spatial integration from texture density was the primary factor in shape perception from kinetic occlusion.

In Andersen and Ni's study [72], they examined conditions similar to those examined by Andersen and Cortese [91]. Subjects were shown kinetic occlusion displays and were asked to identify the 2D shape of a moving form. Information for spatial and temporal integration was manipulated by changing velocity and texture density. Information for spatial and temporal integration was manipulated by changing the texture density of the display. Specifically, an increase in density will result in an increase in spatial and temporal information for the edge boundary. Information for temporal integration was manipulated by varying the speed of motion. Greater velocity of the object increases the rate of accretion and deletion, resulting in an increase in information for temporal integration. If age-related decrements in visual processing are due to changes in temporal integration, then it should be found that object identification performance will be significantly worse for older observers as compared to younger observers at slower speeds.

To isolate the role of spatial integration, they examined variations in performance when the rate of occlusion events remained constant. Specifically, if the velocity of the moving object is decreased and the texture density is increased proportionately, then the spatial separation of accretion and deletion of texture that defines the object boundary is decreased. If age-related decrements in visual processing are due to changes in spatial integration, then object identification performance will be significantly worse for older observers as compared to younger observers at lower texture densities under constant rates of occlusion.

Two experiments were conducted by Ander and Ni [72] to test their hypothesis. In the first experiment, they examined age-related decrements in the perception of 2D shape from kinetic occlusion. The displays consisted of random dots projected onto either a 2D shape (circle, square, diamond, or 5 point star) or the background. The primary variables of interest were dot density, velocity of object, and the absence/presence of background texture (i.e., providing kinetic occlusion information when background texture was present). If older observers have a decreased capacity for spatially integrating kinetic occlusion information, then poorer performance is expected for older observers with a decrease in dot density when background texture is present, particularly when the rate of occlusion is constant.

Older and younger observers showed similar levels of performance for the kinetic occlusion absent condition. When kinetic occlusion was added, both groups showed greater shape identification performance, but a greater increase in performance was found for the younger as compared to older observers. This result suggests that younger observers, as compared to older observers, had a greater capacity for integrating spatial and temporal information for identifying the boundaries of the object.

Based on these results, they conducted additional analyses of the data specifically for the condition when kinetic occlusion was present. No significant interactions were obtained between age and velocity, age and size, or any other higher order interactions with age. It was shown that age-related decrements in shape identification were approximately constant across variations in velocity. The lack of a significant interaction of age and velocity indicates that the age-related decrements are not due to a result of an age-related decline in temporal integration.

Andersen and Ni further examined the conditions in which two pairs of velocity/density combinations have identical spatial/temporal information. They found significant age effects between those conditions, such as between the combination of 1.2 deg/s velocity with 1.2 dots/deg^2 density and the combination of 0.6 deg/s velocity with 2.44 dots/deg^2 density. The interaction of age and performance for these two conditions was significant, indicating significant age effects when spatial/temporal information is constant. The greater age-related differences at lower density conditions indicate age-related decrements in spatial integration. Younger observers had a 32% increase in detection performance when kinetic occlusion information was present, as compared to an 11.5% increase for older observers.

To provide converging evidence in supporting spatial integration hypothesis, they conducted a second experiment. In this experiment, they managed to isolate the effects of spatial integration by manipulating limited lifetime of the stimuli, indicating the duration of individual dots was restricted. When limited lifetime displays are used, the object and background dots in the display would appear for a limited period of time before being randomly repositioned in the display. As a result, the salience of the edge boundary is considerably reduced because the disappearance/appearance of the dots can be either caused by occlusion or simply by the limited lifetime. In order to recover the boundary, the visual system must determine regions of local motion of limited duration and regions of no motion of limited duration. As a result, variations in the lifetime of the dots provide an opportunity to more directly assess the integration of information over time.

If the age-related decrements in using kinetic occlusion information were due to a loss in temporal integration, then decreasing the lifetime of local dot motion should greatly limit the temporal integration, resulting in a subsequent decrease in performance for older observers. Thus in their second experiment, the presentations of individual dots were shown for durations of 16, 33, 66, or 100 ms based on presentations of 2, 3, 4, or 6 frames. The proportion of dots repositioned from frame to frame varied as a function of lifetime. For example, for the two frame lifetime stimuli, 50% of the dots were repositioned on each frame; for the three frame lifetime stimuli, 33% were repositioned, etc.

The result showed a significant main effect of duration, indicating that an increase in lifetime result in increased detection performance. Surprisingly, planned comparisons of the effect of age at each duration condition showed no significant effects. These results suggest that age-related changes in visual processing are the result of changes in spatial integration.

An important question is that whether other aspects of visual processing, which are known to change with age (e.g., motion perception, pattern recognition, and face recognition), may be due to decreased spatial integration. Age-related changes have also been found in visual attention tasks that involve spatial information [94, 95]. Spatial integration may not only explain age-related effects in visual perception, but also be an important factor in age-related changes in performing driving-related visual tasks.

Future Directions

The results of the preset study suggest two practical applications in improving driving safety for older drivers. The first application concerns driving assisting system. Major automobile manufacturers have developed various intelligent safety systems, e.g., "Crash Mitigation Braking System" from Honda, Volvo's "City Safety," Toyota's "Pre-crash System," and Mercedes Benz's "Distronic Plus" system. However, there are two major concerns regarding the current assisting system prototypes. First, any system utilizing optical information, e.g., laser radar, is prone to weather change such as rain and fog. Second, most of these systems were tested on younger drivers which do not accommodate the specific characteristics of older driver population. A successful driving assisting system should take age-related differences in driving

performance into account, as well as the environmental changes in inclement weather conditions.

A second application concerns the potential improvement in driving performance through training for older drivers. Andersen et al. [96] found that following training with subthreshold stimuli, older observers were able to improve performance in a texture discrimination task, and that this improvement was retained for up to 3 months. Indeed, the training, which occurred over a 2 day period, made the older observers to perform equally well as compared to the pre-training performance of college age. In another study, Richards et al. [97] found that older observers could improve their performance in divided attention through repeated training with the UFOV (useful field of view) task. An interesting topic for future research would be to investigate whether older observers can improve driving performance under low visibility conditions using different training techniques.

Bibliography

1. Langford J, Koppel S (2006) Epidemiology of older driver crashes – identifying older driver risk factors and exposure patterns. Transp Res F Traffic Psychol Behav 9:309–321
2. Evans L (2004) Traffic safety. Science Serving Society, Bloomfield Hills, p 179
3. Owsley C, Ball K, Sloane ME, Roenker DL, Bruni JR (1991) Visual/cognitive correlates of vehicle accidents in older drivers. Psychol Aging 6:403–415
4. National Highway Traffic Safety Administration (2008) Traffic safety fact – 2007 data (NHTSA's national center for statistics and analysis. Publication No.DOT HS 810 992). Washington, DC
5. Owsley C, McGwin G Jr, Phillips JM, McNeal SF, Stalvey BT (2004) Impact of an educational program on the safety of high-risk, visually impaired, older drivers. Am J Prev Med 26:222–229
6. Wood JM, Owens DA (2005) Standard measures of visual acuity do not predict drivers' recognition performance under day or night conditions. Optom Vis Sci 82:698–705
7. Ball K, Owsley C, Beard B (1990) Clinical visual perimetry underestimates peripheral field problems in older adults. Clin Vis Sci 5:113–125
8. McGwin G, Chapman V, Owsley C (2000) Visual risk factors for driving difficulty among older driver. Accid Anal Prev 32(6):735–744
9. Glass JM (2007) Visual function and cognitive aging: differential role of contrast sensitivity in verbal versus spatial tasks. Psychol Aging 22:233–238
10. Schachar RA (2006) Effect of change in central lens thickness and lens shape on age-related decline in accommodation. J Cataract Refract Surg 32:1897–1898
11. Richards OW (1977) Effects of luminance and contrast on visual acuity, ages 16 to 90 years. Am J Optom Physiol Opt 54:178–184
12. Derefeldt G, Lennerstrand G, Lundh B (1979) Age variations in normal human contrast sensitivity. Acta Ophthalmol 57: 679–690
13. Owsley C, Sekuer R, Siemsen D (1983) Contrast sensitivity throughout adulthood. Vis Res 23:689–699
14. McFarland RA, Domey RG, Warren AB, Ward DC (1960) Dark adaptation as a function of age: I. A statistical analysis. J Gerontol 15:149–154
15. Domey RG, McFarland RA, Chadwick E (1960) Dark adaptation as a function of age and time. II. A derivation. J Gerontol 15:267–279
16. Chapanis A (1950) Relationships between age, visual acuity and color vision. Hum Biol 22:1–33
17. Kahn HA, Leibowitz HM, Ganley JP, Kini MM, Colton T, Nickerson RS et al (1977) The Framingham Eye Study I. Outline and major prevalence findings. Am J Epidemiol 106:17–32
18. Sekuler R, Hutman LP, Owsley CJ (1980) Human aging and spatial vision. Science 209:1255–1256
19. Long GM, Crambert RF (1990) The nature and basis of age-related changes in dynamic visual acuity. Psychol Aging 5:138–143
20. Trick GL, Silverman SE (1991) Visual sensitivity to motion: age-related changes and deficits in senile dementia of the Alzheimer type. Neurology 41:1437–1440
21. Gilmore GC, Wenk HE, Naylor LA, Stuve TA (1992) Motion perception and aging. Psychol Aging 7:654–660
22. Andersen GJ, Atchley P (1995) Age-related differences in the detection of three dimensional surfaces from optic flow. Psychol Aging 10:650–658
23. Betts LR, Taylor CP, Sekuler AB, Bennett PJ (2005) Aging reduces center-surround antagonism in visual motion processing. Neuron 45:361–366
24. Bennett PJ, Sekuler R, Sekuler AB (2007) The effects of aging on motion detection and direction identification. Vis Res 47:799–809
25. Andersen GJ, Cisneros J, Atchley P, Saidpour A (1999) Speed, size, and edge-rate information for the detection of collision events. J Exp Psychol Hum Percept Perform 25:256–269
26. Andersen GJ, Enriquez A (2006) Aging and the detection of observer and moving object collisions. Psychol Aging 21: 74–85
27. Norman FJ, Clayton AM, Shular CF, Thompson SR (2004) Aging and the perception of depth and shape from motion parallax. Psychol Aging 19:506–514
28. Norman FJ, Crabtree CE, Hermann D (2006) Aging and the perception of 3-D shape from kinetic patterns of binocular disparity. Percept Psychophys 68:94–101
29. Folk CL, Hoyer WJ (1992) Aging and shifts of visual spatial attention. Psychol Aging 7:453–465
30. Kramer AF, Hahn S, Irwin DE, Theeuwes J (1999) Attentional capture and aging: Implications for visual search performance and oculomotor control. Psychol Aging 14:135–154

31. Hartley AA, Little DM (1999) Age-related differences and similarities in dual-task interference. J Exp Psychol 128:416–449
32. Scialfa CT, Thomas DM, Joffe KM (1994) Age differences in the useful field of view: An eye movement analysis. Optom Vis Sci 71:736–742
33. Sekuler AB, Bennett PJ, Mamelak M (2000) Effects of aging on the useful field of view. Exp Aging Res 26:103–120
34. Owsley C, Sloane ME (1990) Vision and aging. In: Nebes RD, Corkin S (eds) Handbook of neuropsychology, vol 4. Elsevier Science, New York, pp 229–249
35. Hoffman L, McDowd JM, Atchely P, Dubinsky R (2005) The role of visual attention in predicting driving impairment in older adults. Psychol Aging 20:610–622
36. Salthouse TA, Somberg BL (1982) Time-accuracy relationships in young and old adults. J Gerontol 37:49–53
37. Salthouse TA, Somberg BL (1982) Isolating the age deficit in speeded performance. J Gerontol 37:59–63
38. Michon JA (1989) Explanatory pitfalls and rule-based driver models. Accid Anal Prev 21(4):341–353
39. Cutting JE (1996) Wayfinding from multiple sources of local information in retinal flow. J Exp Psychol Hum Percept Perform 22(5):1299–1313
40. Wilkniss SM, Jones MG, Korol DL, Gold PE, Manning CA (1997) Age-related differences in an ecologically based study of route learning. Psychol Aging 12(2):372–375
41. DeLucia PR, Mather RD (2006) Motion extrapolation of car-following scenes in younger and older drivers. J Hum Factors Ergon Soc 48(4):666–674
42. Macuga KL, Beall AC, Kelly JW, Smith RS, Loomis JM (2007) Changing lanes: inertial cues and explicit path information facilitate steering performance when visual feedback is removed. Exp Brain Res 178:141–150
43. Chandler RE, Herman R, Montroll EW (1958) Traffic dynamics: studies in car following. Oper Res 6:165–184
44. Brackstone M, McDonald M (1999) Car-following: a historical review. Transp Res Rec F 2:181–196
45. Andersen GJ, Sauer CW (2007) Optical information for car following: the driving by visual angle (DVA) model. Hum Factors 49:878–896
46. Andersen GJ, Sauer CW (2005) Visual information for car following by drivers: the role of scene information. Transp Res Rec F 1899:104–109
47. Ni R, Kang J, Andersen GJ (2010) Age-related declines in car following performance under simulated fog conditions. Accid Anal Prev 42(3):818–826
48. Bian Z, Andersen GJ (2008) Aging and the perceptual organization of 3-D scenes. Psychol Aging 23:342–352
49. Scialfa CT, Guzy LT, Leibowitz HW, Garvey PM et al (1991) Age differences in estimating vehicle velocity. Psychol Aging 6(1):60–66
50. Massie DL, Campbell KL, Williams AF (1995) Traffic accident involvement rates by driver age and gender. Accid Anal Prev 27:73–87
51. Stutts JC, Martell C (1992) Older driver population and crash involvement trends 1974–1988. Accid Anal Prev 24:317–327
52. McGwin G, Brown DB (1999) Characteristics of traffic crashes among young, middle-aged and older drivers. Accid Anal Prev 31:181–198
53. Kang JJ, Ni R, Andersen GJ (2008) The effects of reduced visibility from fog on car following performance. J Transp Res Board 2069:9–15
54. Broughton KLM, Switzer F, Scott D (2007) Car following decisions under three visibility conditions and two speeds tested with a driving simulator. Accid Anal Prev 39:106–116
55. Andersen GJ, Braunstein ML (1998) The perception of depth and slant from texture in 3D scenes. Perception 27:1087–1106
56. Larish JF, Flach JM (1990) Sources of information useful for perception of speed of rectilinear motion. J Exp Psychol Hum Percept Perform 16:295–302
57. Andersen GJ, Cisneros J, Saidpour A, Atchley P (2000) Age-related differences in collision detection during deceleration. Psychol Aging 15:241–252
58. Andersen GJ, Sauer CW (2004) Optical information for collision detection during deceleration. In: Hecht H, Kaiser M (eds) Theories of time to contact, advances in psychology series, North Holland. Elsevier, Amsterdam, pp 93–108
59. DeLucia PR, Bleckley MK, Meyer LE, Bush JM (2003) Judgments about collisions in younger and older drivers. Transp Res part F Traffic Psychol Behav 6:63–80
60. Al-Ghamdi AS (2007) Experimental evaluation of fog warning system. Accid Anal Prev 39:1065–1072
61. Bian Z, Ni R, Guindon A, Andersen GJ (2009) Aging and the detection of collision events in fog. In: Proceedings of the fifth international driving symposium on human factors in driver assessment, training and vehicle design, pp 69–75
62. Ni R, Andersen GJ (2008) Detection of collision events on curved trajectories. Percept Psychophys 70(7):1314–1324
63. Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York
64. Pelli DG, Robson JG, Wilkins AJ (1988) Designing a new letter chart for measuring contrast sensitivity. Clin Vis Neurosci 2:187–199
65. Braunstein ML, Andersen GJ, Rouse MW, Tittle JS (1986) Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. Percept Psychophys 40:216–224
66. Gibson T (1979) The ecological approach to visual perception. Houghton Mifflin, Boston
67. Gibson T (1966) The senses considered as perceptual systems. Houghton Mifflin, Boston
68. Warren WH, Morris M, Kalish M (1988) Perception of translational heading from optical flow. J Exp Psychol Hum Percept Perform 14:646–660
69. Warren WH, Mestre DR, Blackwell AW, Morris MW (1991) Perception of circular heading from optical flow. J Exp Psychol Hum Percept Perform 17:28–43
70. Andersen GJ, Braunstein ML (1985) Induced self-motion in central vision. J Exp Psychol Hum Percept Perform 11:122–132
71. Nguyen B, Zhuo Y, Ni R (2011) Aging and the steering control under reduced visibility conditions. In: Proceedings of the

sixth international driving symposium on human factors in driver assessment, training and vehicle design (in press)
72. Andersen GJ, Ni R (2008) Aging and visual processing: declines in spatial not temporal integration. Vis Res 48:109–118
73. Ni R, Andersen GJ, Rizzo M (2005) Age related decrements in steering control: the effects of landmark and optical flow information. In: Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design, pp 144–150
74. Ni R, Bian Z, Andersen GJ (2009) Age-related differences in using optical flow and landmark information for steering control. In: Proceedings of the IEEE intelligent vehicles symposium, 2009, pp 691–694
75. Ni R, Andersen GJ (2006) Age related performance in driving: steering and collision detection, vision in vehicles 11. Dublin, Ireland
76. Barlow HB (1953) Summation and inhibition in the frog's retina. J Physiol 119:69–88
77. Brown HA, Allison DD, Samonds JM, Bonds AB (2003) Nonlocal origin of response suppression from stimulation outside the classical receptive field in area 17 of the cat. Vis Neurosci 20:85–96
78. Das A, Gilbert CD (1999) Topography of contectural modulations mediated by short-range interactions in primary visual cortex. Nature 399:655–661
79. Stettler DD, Das A, Bennett J, Gilbert CD (2000) Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron 36:739–750
80. Bullier J, Hupe JM, James AC, Girard P (2001) The role of feedback connections in shaping the responses of visual cortical neurons. Prog Brain Res 134:193–204
81. Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation of frontal cortex. Nature 421:370–373
82. Roberts MJ, Zinke W, Guio K, Robertson R, McDonald JS, Thiele A (2005) Acetylcholine kinetically controls spatial integration in Marmoset primary visual cortex. J Neurophysiol 93:2062–2072
83. Ledgeway T, Hess RF (2002) Rules for combining the outputs of local motion detectors to define simple contours. Vis Res 42:653–659
84. Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local 'association field'. Vis Res 33:173–193
85. Kovacs I (1996) Gestalten of today: early processing of visual contours and surfaces. Behav Brain Res 82:1–11
86. Aspell JE, Wattam-Bell J, Braddick O (2006) Interaction of spatial and temporal integration in global form processing. Vis Res 46:2834–2841
87. Bair W, Movshon JA (2004) Adaptive temporal integration in direction-selection neurons in Macaque visual cortex. J Neurosci 24:7305–7323
88. Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77
89. Baker CL, Braddick OJ (1985) Temporal properties of the short-range process in apparent motion. Perception 14:181–192
90. Andersen GJ, Cortese JM (1989) 2-D contour perception resulting from kinetic occlusion. Percept Psychophysic 46:49–55
91. Kellman PJ, Shipley TF (1992) Perceiving objects across gaps in space and time. Curr Dir Cogn Sci 1:193–199
92. Lee S-H, Blake R (1999) Visual form created solely from temporal structure. Science 284:1165–1168
93. Blake R, Lee S-H (2005) The role of temporal structure in human vision. Behav Cogn Neurosci Rev 4:21–42
94. Thorton WJL, Raz N (2006) Aging and the role of working memory resources in visuospatial attention. Aging Neuropsychol Cogn 13:36–61
95. Greenwood PM, Parasuraman R (2004) The scaling in visual search and its modification in healthy aging. Percept Psychophys 66:3–22
96. Andersen GJ, Ni R, Bower JD, Watanabe T (2010) Perceptual learning, aging, and improved visual performance in early stages of visual processing. J Vis 10(13):4, 1–13
97. Richards E, Bennett PJ, Sekuler AB (2006) Age related differences in learning with the useful field of view. Vis Res 46:4217–4231

Drug Discovery in Ocean

David J. Newman, Gordon M. Cragg, Paul G. Grothaus
Natural Products Branch Developmental Therapeutics Program, National Cancer Institute, NCI-Frederick, Frederick, MD, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Antitumor Agents
Anti-infective Agents
Zinconotide, Cone Snail Toxin
Future Directions
Bibliography

Glossary

Biodiscovery The investigation of natural products (secondary metabolites) from organisms with the aim of determining if they have biological activities of utility in human diseases.

Derivatized A chemical structure that resembles a known published structure in general but has specific modifications made by nature or man.

Dredging collections Use of towed sledges or nets designed to skim the sea bed/ledges horizontally or at an angle, with organisms being separated at the surface.

Manned submersible A self-propelled small submarine equipped with collection devices that usually is used in the depth range of 50–1,000 m when used for biodiscovery.

Orthotopic tumor A tumor model where the human (or animal) tumor is implanted in its original site rather than in a subcutaneous or intraperitoneal site distinct from its site of nominal origin. An example would be the implanting of a brain-derived tumor into the cranial cavity of the animal instead of into the flank or peritoneum.

Proteasome inhibitor A chemical compound that blocks the protein degradation by cell organelle.

Remotely operated vehicle (ROV) A torpedo-shaped self-propelled underwater vehicle controlled via a tether from a mother ship, or autonomously via an internal guidance device. Collection baskets and video/tv recording devices are part of the general equipment.

Secondary metabolites Chemical compounds encoded by genes and produced by organisms that are not required for the basic life processes of the producer. These are usually considered to be produced when the organism needs to defend itself from attack, or wishes to attack another organism to gain an advantage.

Sessile organism A normally nonmotile marine organism that requires a "foothold" on a substrate in order to survive and grow. Almost all are filter feeders, moving large amounts of seawater through their bodies on a daily basis. Marine sponges (phylum Porifera) are prime examples of this type of organism.

Shallow water collections Usually considered to be from simple wading to the "regular" limit of SCUBA (self-contained underwater breathing apparatus) which is <50 m for well-trained and experienced divers.

Definition of the Subject

The oceans of the world cover roughly 71% of the planet and have a median depth of $>3{,}000$ m and a mean depth of 3,800 m. Of this vast expanse, less than 5% of the deep sea has been explored in any way and less than 0.01% of the deep sea floor has been sampled in detail [1]. What is often not realized is that, of the 36 animal phyla listed taxonomically, 34 are found in the marine environment [2] and of these, approximately half are only found in marine environments. Of the remainder, another approximate half are both marine and terrestrial (though predominately marine) and only one, the phylum Onychophora (the velvet worms) is exclusively terrestrial.

Although organisms had been recovered from depths in the 1800s and obviously had been seen for many centuries on coral reefs in shallow waters and beaches, it was not until the Challenger expedition in the 1872–6 time frame, and then the Galathea expedition in 1950–2 that collected live animals from the Philippines Trench at 10,190 M [3] that it was demonstrated that, with the possible exception of the anoxic zones in the Black Sea, living animals could be found at all depths. The reason for using "possible" as a modifier above is that the formaniferans tube worms found at "black smokers" are in anoxic environments and utilize microbes of the sulfur cycle to survive and grow.

Who first decided to investigate the marine environment as a source of medicaments is unknown, but the Japanese and Chinese herbals do contain various mixtures that were used as part of traditional medicines, and certainly, the toxic properties of marine-derived products were well-known centuries ago, with an example being the puffer fish (Fugu) in Japanese cuisine. From a historical perspective, "Tyrian Purple," a dyestuff from a Mediterranean mollusk used in Roman times, actually has significant activity in some cancers and was used as a part of traditional Chinese medicine (TCM) for the treatment of leukemia. It is only recently that it was realized that the active chemical compounds (marine sourced and plant sourced, respectively) were of the same basic chemical class.

Thus, this article will deal with the potential of the marine (aquatic) environment and its organisms as leads to agents of value to humans either directly or indirectly. What will become obvious as the stories unfold is that the original organism (usually a marine invertebrate) may or may not be the actual producer of the metabolite(s) of interest. Where the true producer is known, it will be identified, but in a significant

number of cases, the evidence is still circumstantial, though in other cases, the organism is directly identifiable.

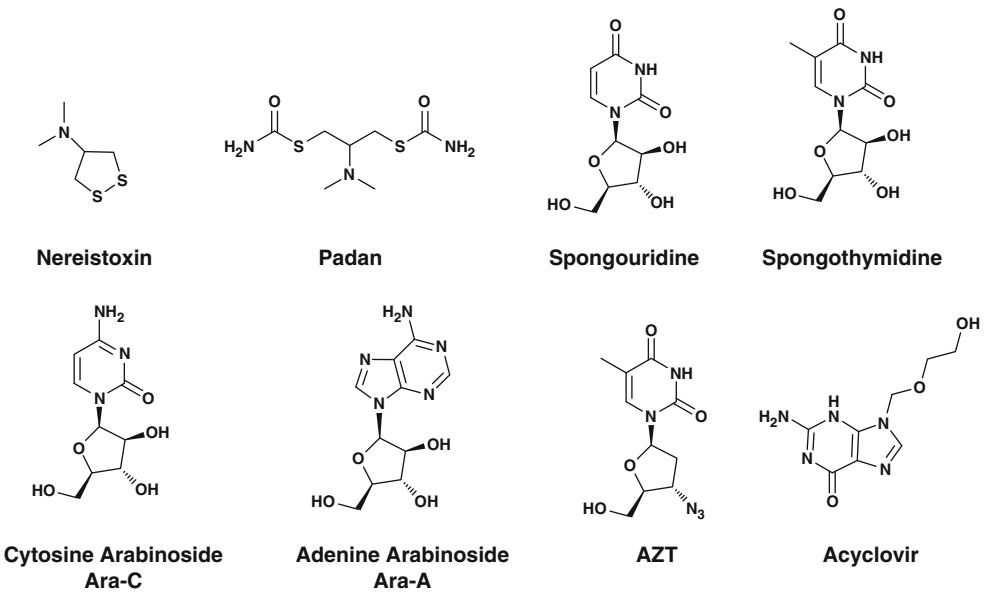
Introduction

Though found in both marine and aquatic environments ranging from the Sea of Japan to Lake Victoria in Central Africa, one can argue that one of the first generally used "marine" invertebrate-sourced secondary metabolites was the toxin known as nereistoxin (Fig. 1) from the flatworm *Lumbrineris brevicirra* (also named *Lumbrineris heteropoda*). This worm had probably been used for centuries by fishermen to stun fish within the lake, thus permitting them to harvest their catches rather easily. One might almost call it a "chemical dynamite fishing system" by analogy to the explosive methods used by fishermen in countries such as the Philippines and other areas of SE Asia. In addition, Japanese fishermen had known for years that the flatworm had anti-insecticidal properties as it killed carnivorous insects when they landed on it.

In 1934, Nitta isolated the active principle in the flatworm with a structure being proposed by Okaichi and Hashimoto in 1962 and confirmed by total synthesis reported by Hagiwara et al. in 1965 [4]. This compound was shown to be an acetylcholine antagonist, and over the next few years, insecticides were developed from the base structure, with a close analog, Padan® (Fig. 1) being marketed by Takeda in 1967.

As mentioned earlier, there had been sporadic reports from at least two millennia ago of pharmacologically active agents being isolated and identified from marine organisms, but it was not until the twentieth century that any form of systematic investigation occurred. Thus in the early 1960s to early 1970s, academic groups in the USA and pharmaceutical company-linked organizations such as the Roche Institute for Marine Pharmacology in Australia reported their findings in a variety of formats. For information on these and earlier studies, one should consult the 1976 review by Ruggieri [5].

One may argue, and the authors and others have done so on a number of occasions [6, 7], that the seminal discoveries, and thus the impetus for the investigation of marine biodiversity and the subsequent vision of marine-derived drugs on the market, can be traced to identification by Bergmann in the early 1950s of the arabinose-containing bioactive nucleosides spongothymidine and spongouridine (Fig. 1) from



Drug Discovery in Ocean. Figure1
Early agents from marine sources and their subsequent derivatives

the Caribbean sponge *Tethya crypta* [8–10]. These discoveries overthrew the then current dogma that a nucleoside had to have either ribose or deoxyribose as the sugar moiety in order to demonstrate biological activity.

The subsequent explosion of compounds is described in the reviews by Suckling [6] and Newman et al. [7]. These discoveries led to the identification of a close analog, cytosine arabinoside (Fig. 1) as a potent antileukemic agent; this compound was subsequently commercialized by Upjohn (now Pfizer) as Ara-C. Other closely related compounds such as adenine arabinoside (Ara-A) (Fig. 1), an antiviral compound synthesized and commercialized by Burroughs Wellcome (now Glaxo SmithKline), and later found in the Mediterranean gorgonian *Eunicella cavolini*. Similar reasoning led to investigations on the minimum size of sugar rings and other substitutions on the sugar ring[s], thus azidothymidine (AZT) (Fig. 1) and even acyclovir (Fig. 1) can be traced back to this initial discovery.

In the subsequent sections, the discussion will be separated on the basis of pharmacologic activities rather than by nominal source since, as mentioned earlier, it is now being realized that what was thought to be the product of a marine invertebrate may well be produced by a microbe, a consortium of microbes, and/or interactions between microbe(s) and invertebrate host. Where data exist as to actual or probable source, the information will be cited.

Antitumor Agents

Introduction

By the early 1960s, the terrestrial plant world had been explored and reported by a number of groups from the 1940s onwards, with agents such as the Vinca alkaloids first being reported as potential antitumor agents in the late 1950s. Similarly, the terrestrial microbial world had been explored in a systematic fashion from roughly the late 1940s with success in both antibiotic and antitumor areas, and a number of agents still in current clinical use were first identified in the late 1950s to early 1960s in industry.

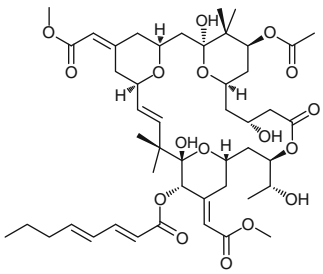
Similarly, although there had been a number of reports of biologically active agents from marine (and freshwater) environments (cf. discussion on arabinosides above), predominately by academic researchers who received funding from either NIH institutes or NSF for basic research, no systematic explorations had been performed on marine environments as sources of medicaments.

Starting in the early 1960s, the National Cancer Institute (NCI) the largest institute within the US government's National Institutes of Health (NIH), expanded its horizons beyond the testing of synthetic compounds as experimental antitumor agents and began large-scale studies of natural products. Plant materials were collected and recollected when required (and if feasible) in conjunction with the USDA. Microbial products were usually provided by pharmaceutical companies, and small numbers of marine invertebrates were also obtained by purchase from collectors.

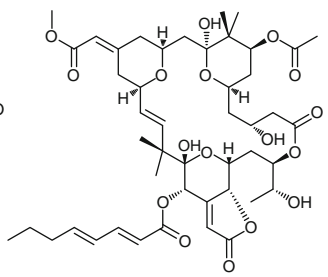
These materials were then extracted by predominately academic groups; the extracts tested initially in whole animal models (mouse and rat tumors) and later in fast-growing animal leukemia cell lines by US government contract laboratories with "active materials" then being isolated and identified by academic collaborators. At the same time, academic investigators including those who were also acting as NCI collaborators were also expanding their investigator-initiated grant applications towards the marine environment with the aim of collecting and utilizing marine invertebrates as sources of drug leads (marine biodiscovery).

Early Marine-Sourced Antitumor Compounds

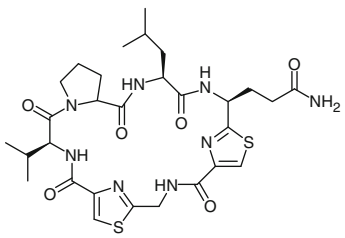
Introduction The two antitumor compounds that came from what might be considered as the earliest systematic investigations of the marine environment for such agents, as distinct from finding activities for isolated compounds (that might be considered as the marine equivalent of phytochemical investigations), were two agents reported by the Pettit group at Arizona State University working with the NCI. These were bryostatin 3 and dolastatin 10 (Fig. 2). The former was isolated from the well-known fouling organism *Bugula neritina* and the latter from the nudibranch *Dolabella auricularia*. Subsequently, each was found to be a representative of a family of somewhat similar compounds, examples of each going into extensive isolation/synthesis programs and then clinical trials in man over the next 20 plus years. The earlier history of both the dolastatins [11] and the bryostatins [12] was



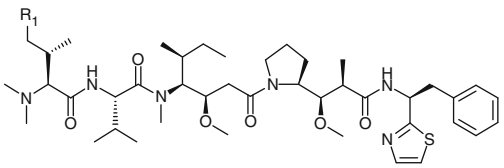
Bryostatin 1



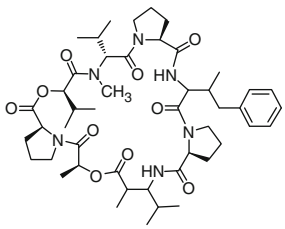
Bryostatin 3



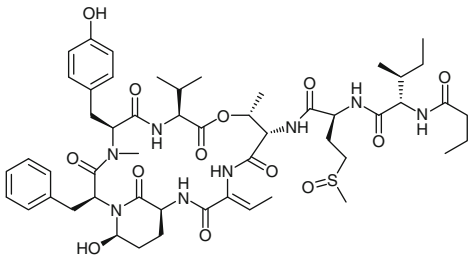
Dolastatin 3



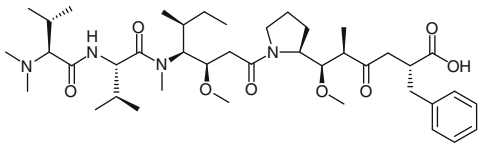
Dolastatin 10; R1 = H
Symplostatin 1; R1 = CH3



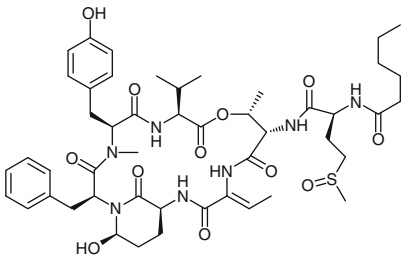
Dolastatin 16



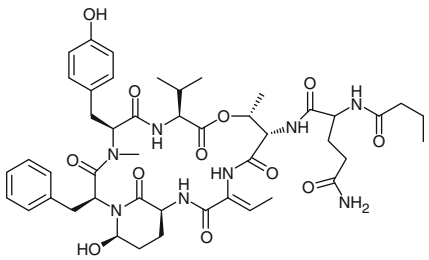
Symplostatin 2



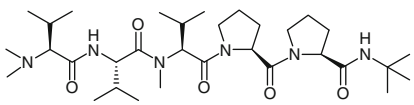
Symplostatin 3



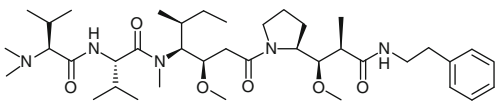
Somamide A



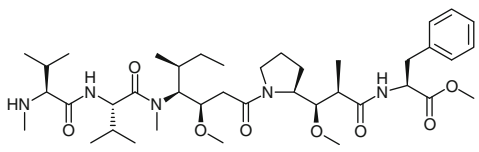
Somamide B



Tasidotin



Auristatin PE



Monomethylauristatin F

Drug Discovery in Ocean. Figure 2
Early antitumor agents from marine sources and chemical derivatives

reported in detail in the 2005 book *Anticancer Agents from Natural Products*. At that time, both classes of agents had been in multiple trials with only marginal success.

Bryostatins 1–20 and Bryologs In the case of the bryostatins, over the roughly 35 years since the initial report of bryostatin 3 in 1970 to the review in 2005, 19 variations on the bryostatin structure had been reported with syntheses of at least three in that time frame. From the early 1980s, bryostatin 1 (Fig. 2), which was the most abundant of the group and had been isolated and purified to cGMP quality mainly by workers at the NCI-Frederick in the late 1980s, and on a small scale by Pettit, went into a multiplicity of clinical trials (at phases I and II) as both a single agent and in conjunction with cytotoxins. However, to date, only a few patients showed responses to the agent in spite of being in over 80 clinical trials in this time frame. Currently, there are just two trials listed as of February 2011 that are still active. One phase II, studying paclitaxel and bryostatin 1 is active but not recruiting patients, and the other, a phase I trial with temsirolimus and bryostatin 1 that is listed as active and recruiting. What is of interest, however, is that the base molecule is now being investigated as a potential treatment for Alzheimer's disease, with a phase I trial approved in 2008 under the sponsorship of the Blanchette Rockefeller Neurosciences Institute in West Virginia, but the trial is not yet open for participants.

There have been some quite recent reviews and reports on the earlier syntheses of bryostatin and the simplified versions made by Wender that are known as the "bryologs," together with the research work that has led to the possibilities for the treatment of Alzheimer's disease referred to above. The interested reader should consult the excellent 2010 review article by Hale and Manaviazar [13] for a thorough discussion of all of the chemistry involved in the bryostatins, the papers by the West Virginia neurosciences groups [14–18] with a very interesting potential linkage to Alzheimer's treatment and bryologs recently published by Khan et al. [19].

Finally, the bryostatins may well be produced by an as yet non-cultured commensal microbe or microbes found in the pallial sinus of the larvae of the "nominal" producing organism *Bugula neritina*. As a result of extensive ecological and genomic analyses by Haygood and her collaborators, there is now very good evidence linking biosynthetic gene clusters from this commensal microbial organism (not yet identified; simply given the name "*Candidatus endobugula sertula*") to the formation of bryostatin. The discussion is beyond the scope of this article, but for readers who are interested, they should consult the following research papers that cover the discoveries [20–31]. One may ask whether the effort expended on the bryostain story was worth it? In the view of the authors, definitely! The chemistry, microbiology, studies of tumor systems, and their interactions would not have occurred without molecules such as these. The ultimate aim is to come up with chemical derivatives that will mimic the natural molecule but have better pharmacological properties.

Dolastatins and Derivatives (ILX651 and Auristatin PE) The dolastatins, unlike the bryostatins, had to be synthesized in order to obtain enough material for preclinical and clinical trials. The pre-2005 chemistry related to the synthesis of these compounds was discussed in detail by Flahive and Srirangam [11], though a little earlier, the proof that natural dolastatins and closely related peptides were definitively produced by cyanobacteria, was given in a series of publications. These reported the isolation of dolastatins 3, [32] 10, [33] 12–14 [34, 35], and 16 [36] directly from varieties of the blue-green algae *Lyngbya majuscula* and *Symploca hydnoides*, along with other novel closely related bioactive compounds such as the symplostatins 1–3 [37–39] and somamides A and B (Fig. 2) [35].

Dolastatin 10 was in a significant number of phase I and phase II clinical trials but was not successfully developed beyond that level. There are two "hold over" trials still listed as of Feb 2011 in the NCI's Clinical Trials web site, where it is reported as being studied as a single agent in a phase I study at M.D. Anderson as a treatment for various leukemias, and in a phase II study at the University of Chicago against liver-related carcinomas. No updates have been posted since 2008, nor have there been any publications directly related to the trials, but very recently, Siemann published a review on tumor-vascular disrupting agents that included dolastatin 10 and some close relatives within this mechanism of action category [40] including other agents that also bind to the colchicine-binding site of tubulin.

As a result of the experience with dolastatin syntheses, the Pettit group, and others, synthesized a variety of agents based upon dolastatin 10 and 15. Of these agents, ILX651 or tasidotin (Fig. 2), based upon dolastatin 15, entered clinical trials but currently, no active trials are shown, and the current owners of the compound, Genzyme (now owned by Sanofi-Aventis), reported in 2008 that they will continue investigating it at the preclinical level as an oral antitumor drug.

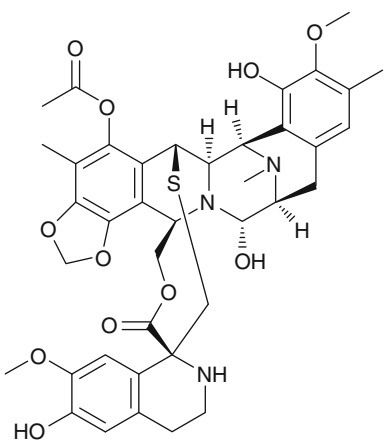
Another derivative, auristatin PE (Fig. 2) also entered clinical trials as both a tubulin interactive agent and also with potential as a vascular disrupting agent [40]. It too, did not demonstrate objective responses [41, 42]. This compound (some with subtle chemical modifications such as monomethyl auristatin E, also known as vedotin), and with or without a small chemical linker, was used as warheads by Seattle Genetics attached to selective monoclonal targeting antibodies. Variations differing in the antibody and the specific linker are currently in phase III (brentuximab vedotin, SGN-35), phase II (glembatumumab vedotin, CDX-011), phase I (PSMA-ADC, MEDI-547 and MN-immunoconjugate) with another (anti-CD19-vcMMAE), and in preclinical trials.

A variant of auristatin PE (known as auristatin F (Fig. 2) in which the C-terminal amino acid was changed to phenylalanine) provided another monoclonal antibody-warhead combination (1 F6-MMAF, SGN-75). This combination is currently in phase I clinical trials.

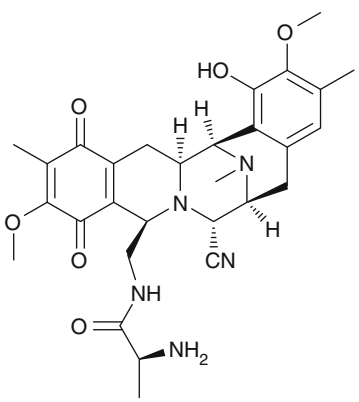
These clever modifications demonstrate how compounds based on marine natural products that exhibit interesting activities, but have pharmacologic properties that hinder clinical development, can be further optimized (rescued?) by clever chemistry and biology, with a new drug application (NDA) being filed by Seattle Genetics for SGN-35 late in 2010 for treatment of Hodgkin's lymphoma.

Antitumor Agents in Use

Trabectedin (Yondelis®, Et-743) The first, and currently only anticancer compound "directly from the sea," in the sense that the molecule approved is identical to the natural product for the treatment of cancer, is trabectedin (Yondelis®) (Fig. 3). This complex tetrahydroisoquinoline alkaloid was approved by the EMEA, the European equivalent of the FDA, in September 2007 for the treatment of sarcoma and was launched in Sweden, Germany and UK in late 2007 for that indication. Subsequently, it was approved for treatment, in conjunction with liposomal doxorubicin, of recurrent ovarian cancers that were sensitive to antitumor drugs containing platinum at the launching in 2009 in Sweden, UK and Spain. Currently, there are



ET743; Trabectedin



Cyanosafracin B

Drug Discovery in Ocean. Figure 3
The first "direct from the sea" antitumor agent and semisynthetic precursor

multiple clinical trials, all are ongoing and recruiting, listed in the NCI clinical trials database (http://clinicaltrials.gov), and details of the clinical trials that led to the ovarian cancer approval have recently been published [43, 44].

This compound belongs to a series of compounds originally reported by two different research groups in 1990 [45, 46]. The compound was recognized as a derivatized member of the well-known saframycin class of terrestrial antibiotics, and it was therefore possible that there might be a microbial component in the production of the compound by the invertebrate *Ecteinascidia turbinata*.

Trabectedin, then known by the abbreviation Et-743, was licensed by the discoverers to the Spanish pharmaceutical company, PharmaMar, and followed a very checkered development path including large-scale wild collections of the source tunicate *Ecteinascidia turbinata*, both in-sea and on-land aquaculture (all of which produced enough harvested materials for the initial and early clinical trials). The compound, though synthesized early in the study by Corey's group, was ultimately produced for later clinical trials by semisynthesis starting with cyanosafracin B (Fig. 3). This intermediate compound was produced by fermentation of a marine-derived *Pseudomonas fluorescens* and additional details of this agent can be found in the 2005 review by Henriquez et al. [47].

In addition to the review cited above, Velasco et al. reported the identification of the producing gene cluster for saframycin from the *P. fluorescens* culture used to produce the intermediate, thus presenting possibilities for combinatorial biosyntheses for trabectedin [48]. In addition to this report, approximately 2 years earlier, potential bacterial sources for trabectedin from Mediterranean Sea-sourced *Ecteinascidia turbinata* were identified by Moss et al. [49] which was followed in 2007 with an examination of the microbial assemblages associated with both Caribbean and Mediterranean isolates of *Ecteinascidia turbinata*.

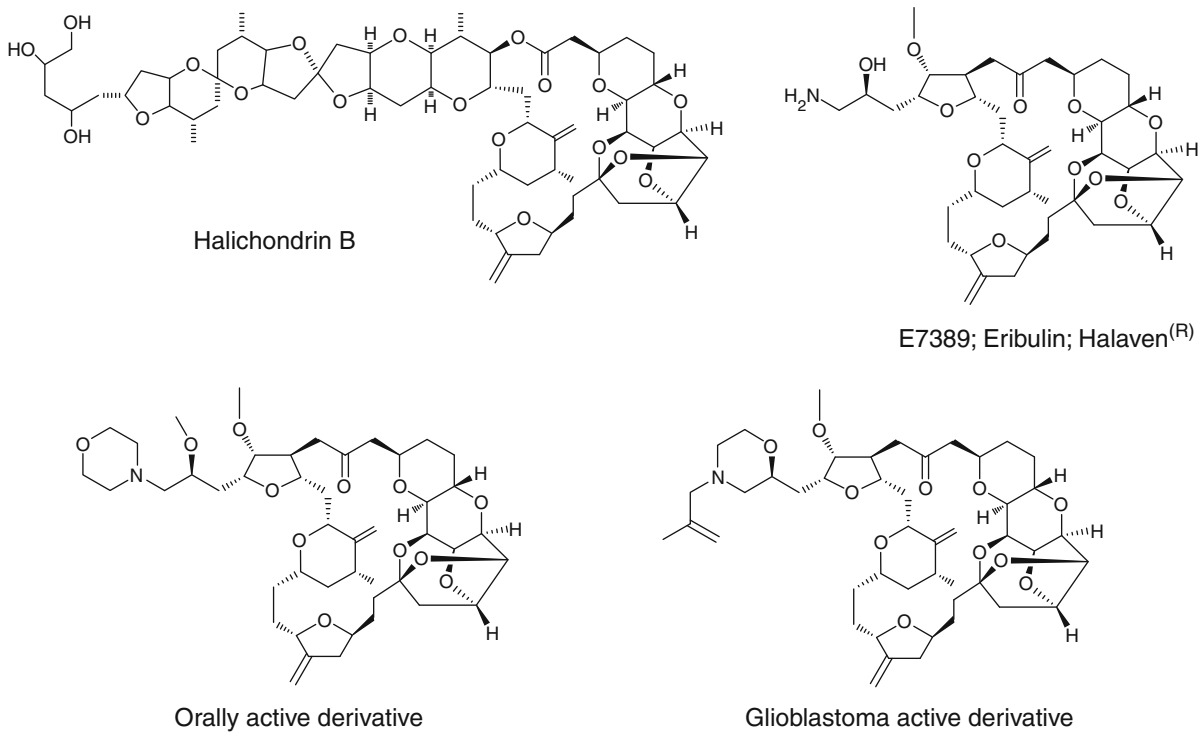
In this report, five potentially persistent bacteria were identified, with one, *Candidatus Endoecteinascidia frumentensis*, occurring in *Ecteinascidia turbinata* in both geographical locations [50]. Although this is not absolute proof, such occurrences are indicative of microbial involvement in the production of trabectedin and require further investigation similar to the research conducted for the putative bryostatin producer (*vide infra*).

From a mechanistic aspect, a number of reports have been published in the literature over the last few years giving possibilities as to the mechanism of action (MOA) of trabectedin when tumor cells are treated in vitro. However, a significant problem with a number of these published reports is that the concentration(s) used in the experiments were frequently orders of magnitude greater than those that demonstrate activity in vitro. Since the "active" levels (in cellulo) are in the low nM to high pM range, care should be taken when evaluating published work on the MOA of this compound.

At physiologically relevant concentrations, the MOA(s) of trabectedin have been shown to include the following: effects on the transcription-coupled nucleotide excision repair process (TC-NER) and interaction between the Et-743 DNA adduct and DNA transcription factors, in particular, the NF-Y factor [51]. Recent pharmacogenomic analyses have identified a series of genes involved in the sensitivity of tumor cells to this agent, and the paper by Jimeno et al. [52] should be consulted in order to see the patterns identified.

In 2007, Soares et al. [53] demonstrated that adducts formed with ET-743 and DNA were stable and could convert into double-strand breaks (DSBs) a number of hours after initial formation. In addition, loss of the homologous recombination repair function, although having no effect on the initial number of DSBs, was associated with persistence of these lesions which could give rise to extensive chromosomal abnormalities, and thus, sensitivity to the drug. In addition to this report, two other papers that demonstrated other more subtle aspects of its cellular interactions have been published; the first in 2006, demonstrating that the demethylated analog ET-729 has a differential affinity for the CGA DNA triplet -binding site compared to ET-743 [54] and the second in 2007, demonstrating that sensitivity correlates with mutated p53 in low passage sarcoma cell lines [55].

Eribulin (Halaven®, E7389) This agent, which is a wholly synthetic molecule modeled on the ring structure of the naturally occurring antitubulin compound halichondrin B (Fig. 4), came from a tour-de-force



Drug Discovery in Ocean. Figure 4
Halichondrin B and the totally synthetic compounds derived from its structure

based upon the synthetic method for halichondrin B first reported by Kishi's group in 1992 [56]. During this synthesis and the investigation by scientists at the Eisai Research Institute (ERI) in Woburn, MA of the synthetic intermediates, Kishi and ERI scientists realized that the active part of the molecule resided in the macrolide ring (approx MW of 600) and not in the "tail" (the remaining 400 + of the overall 1,000 + MW). Chemists at the ERI, working very closely with the Kishi's group, synthesized over 200 molecules and, in conjunction with the Developmental Therapeutics Program (DTP) at NCI, they chose the modified truncated macrocyclic ketone (E7389) (Fig. 4) as the candidate compound when compared in in vitro and in vivo studies to pure halichondrin B (obtained by DTP in conjunction with New Zealand scientists). Much fuller details of the synthetic and base biological information were published by the leaders of the synthetic studies in 2005 [57]. This review article should be read by the interested reader for fuller details of the evolution of this compound through early 2005.

Eribulin, as with the "natural product parent," is a tubulin-interactive agent with very potent activity at the nanomolar level in in vitro studies. In 2005, Jordan et al. [58] reported that suppression of microtubule growth was the primary antimitotic mechanism of E7389 with differential effects due to concentration when studied in MCF7 cells. At low concentrations eribulin, potently inhibited microtubule dynamics resulting in prolonged arrest of mitosis and inducing apoptosis, whereas at tenfold the IC_{50} value (or higher concentrations), it induced depolymerization [58, 59] with extension of the experiments to demonstrate eribulin's interaction with centromere dynamics [60].

Within the same time frame, Jordan's group and others expanded studies to demonstrate that E7389/eribulin and halichondrin B could be modeled into a site close to the vinca site in tubulin [61] which was extended by use of analytical ultracentrifugation experiments by Alday and Correia [62]. Their data strongly suggested that eribulin is a strong inhibitor of the 1:2 stathmin-tubulin complex, and may also be a global inhibitor of tubulin polymer formation as a result of

disruption of tubulin–tubulin contacts at the interdimer interface. Also of interest in this work was the discovery that the penultimate precursor of eribulin (ER-076349), where the terminal amino group is an hydroxyl, is not as potent an inhibitor of the stathmin-tubulin complex formation as eribulin. Though it too, perturbs tubulin–tubulin contacts, it appears to have a more direct effect upon the tubulin polymer.

That the actual interaction of eribulin with tubulin may be even more complex than was originally thought is shown by the recent data reported by Jordan's group in a 2010 paper [63]. By using ^{3}H-eribulin, they reported a very high affinity site (K_d 400 $\pm$ 200 nM) on 25% of the tubulin mass (which might be the $\alpha\beta_{III}$ tubulin dimer) with a stoichiometry of 0.26 $\pm$ 0.12 moles of eribulin per tubulin dimer. Another high affinity site (K_d 3.5 $\pm$ 0.6 μM, stoichiometry of 14.7 $\pm$ 1.3 eribulin-tubulin dimer) was also identified together with a low affinity site (K_d $\sim$ 46 $\pm$ 28 μM, stoichiometry of 1.3 $\pm$ 0.4 saturable sites per tubulin dimer).

The binding isotherm from these results is highly complex which might well be due to competition for eribulin between soluble tubulin and microtubules, though there is photo and electron micrographic evidence in the paper, and in earlier ones from the same group, for binding to the plus end of microtubules and preferentially to β-tubulin. This is concordant with the ultracentrifugation data reported above [62]. As mentioned in the Smith et al. paper [63], further experimentation will be necessary in order to explain these complex results, including the inhibition of eribulin binding to microtubules in vitro by vinblastine when the eribulin concentration is $<$5 μM versus the reverse when $>$5 μM.

In the last 2 years or so, there have been three publications from the Kishi's group detailing improved methods for the synthesis of eribulin and, though written from an academic perspective, these papers consider the relative costs of production [64–66]. In addition, there was a 2007 review by Wang et al. [67] that gave extensive coverage of the patent routes to eribulin, with some duplication of the work presented in the 2005 review by Yu et al. [57]. This was followed in early 2009 by two excellent review articles from the Phillips group at the University of Colorado, covering the published syntheses by many groups of halichondrin B, norhalichondrin B, and E7389 [68, 69].

Very recently, three papers have been published by the Eisai group demonstrating that with only a relatively minor change to the "tail" of the molecule (Fig. 4), the base structure now had a much lower propensity for inducing P-glycoprotein susceptibility, when the diol that is the eribulin precursor was changed to either the dimethoxy substituent or the terminal amino group was changed to a methoxy group. Both compounds were potent in vivo and had a reduction of approximately 30-fold in terms of being substrates for P-glycoprotein compared to eribulin [70].

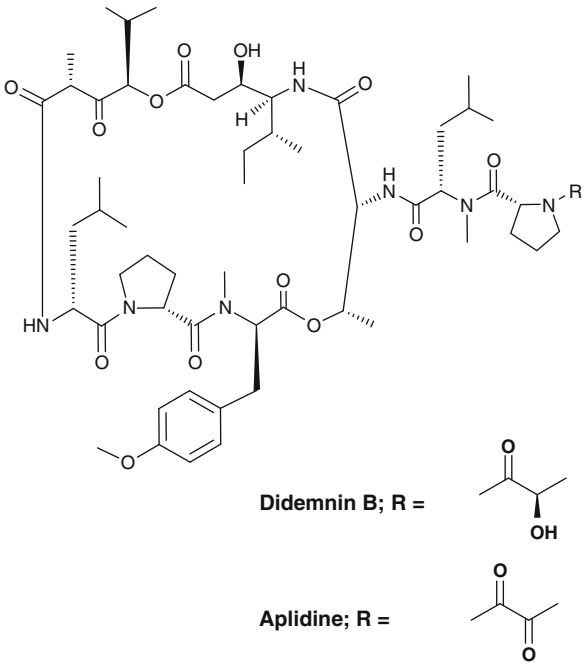
Continuing on with the eribulin studies, the Eisai group then demonstrated that by substituting a morpholine for the terminal amino group and changing the sidechain hydroxyl group to a methoxyl, the compound (Fig. 4) demonstrated oral activity in a subcutaneous LOX melanoma model and maintained a low susceptibility to P-glycoprotein induction [71].

Since there are very few treatments for brain tumors, the group then modified the base eribulin molecule by ring closure at the "tail" to give another morpholino derivative (Fig. 4). This molecule was subtly different from the orally active compound referred to above, and this molecule demonstrated intravenous in vivo activity in an orthotopic murine model of a human glioblastoma [72].

Although it effectively took from the original report on halichondrin B in 1985 until late 2010 to approve eribulin, the interplay of academic, industrial, and government laboratories in three continents led to the novel agent, perhaps the most complex drug molecule yet produced by total synthesis, but what is also of great import is that without the structure of the halichondrin B from a Japanese sponge and the subsequent purification of halichondrin B by NZ scientists and the NCI, eribulin would not have been approved, nor would, the very interesting and novel agents referred to in the preceding three paragraphs, have been synthesized.

Agents in Clinical Trials

Aplidine (Plitidepsin) Aplidin which is formally dehydrodidemnin B (Fig. 5), and thus, a very close chemical relative of the first direct-from-the-sea antitumor compound didemnin B (Fig. 5), was



Drug Discovery in Ocean. Figure 5
Two compounds differing by two hydrogen atoms

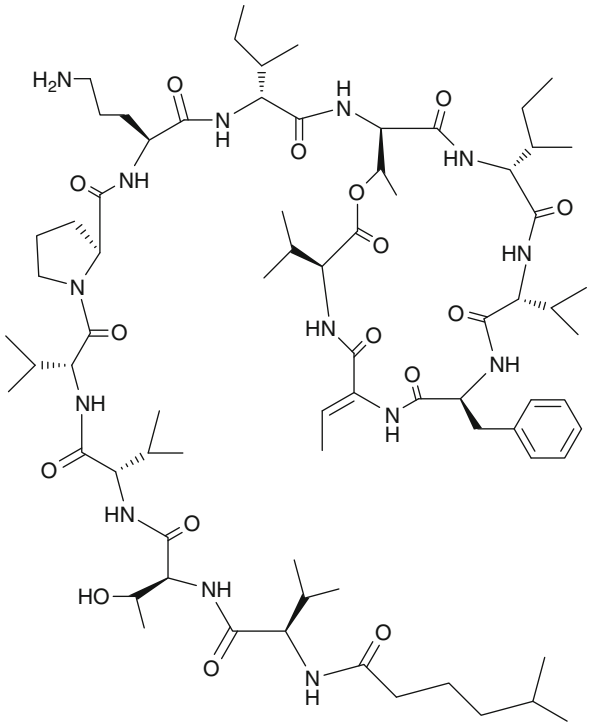
originally isolated from *Aplidium albicans* and first reported in a patent application in 1989, with an UK patent issued in 1990. It was formally identified in the chemical literature in a paper from Rinehart's group in 1996 [73] on the structure-activity relationships among the didemnins. The earlier work on aplidin, its entry into phase I and II trials and the preferred method of synthesis, was described in detail through late 2004 by PharmaMar scientists [47]. At that time, the dose-limiting toxicity was muscle pain responsive to either dose limitation or addition of carnitine. Interestingly, in the presence of carnitine, the maximum tolerated dose could be increased by 40% to 7 mg/m^2. Since those earlier reports, significant numbers of phase I and II trials have been reported and or initiated in the USA and Europe. Currently, this agent is in three clinical trials listed in the NCI clinical trials database, including one phase III trial with or without dexamethasone against resistant multiple myeloma.

The precise MOA of this agent is not yet fully described, but it appears to block the vascular endothelial growth factor (VEGF) secretion and blocks the corresponding VEGF-Receptor-1 (also known as Flt1) autocrine loop in leukemic cells [74]. Using data from a series of ex vivo studies with leukemic blasts from pediatric and adult patients where cells underwent massive apoptosis at levels of 5 nM below the blood levels achievable in man [75], workers at PharmaMar developed a pharmacogenomic model that led to a molecular fingerprint for sensitivity to this agent using the "Oncochip" array. Full details are given in the 2006 paper [52]. Recently, a thorough analysis of the data available to PharmaMar scientists and their collaborators on the MOA of this agent was published, but even today, the exact mechanisms have not been identified, though effects upon the VEGF system are noted, together with the potential involvement of lipid rafts in cellular membranes [76].

What is interesting both chemically and pharmacologically is that the removal of two hydrogen atoms, i.e., conversion of a lactyl side chain in didemnin B to a pyruvyl side chain in aplidin, appears to significantly alter the toxicity profile as this is the only formal change between the two structures. However, the comments on dosage regimens for didemnin B from Vera and Joullie [77] should perhaps be taken into account in any future comparisons.

Kahalalide F This cyclic depsipeptide (Fig. 6) was first isolated as part of a number of similar molecules from the Sacoglossan mollusk, *Elysia rufescens*, after the mollusk grazed on the green macroalga *Bryopsis* sp. After isolation and identification from the invertebrate, it was discovered that the depsipeptide also occurred in the alga, but at a much lower weight basis when compared to the yield from the mollusk. Thus, the invertebrate significantly (20-fold or better) concentrated the depsipeptides [78]. The compound was licensed to PharmaMar by the University of Hawaii in the 1990s, and it entered preclinical testing. Using solid-phase peptide techniques, the compound was synthesized in a very efficient manner by a group in the chemistry department at the University of Barcelona [79], and kahalalide F entered phase I clinical trials in Europe in December 2000 for the treatment of androgen-independent prostate cancer.

A variety of mechanisms have been attributed to this compound. It was known to target lysosomes [80], which implied a selectivity for tumor cells such as prostate tumors that exhibited a high lysosomal



Kahalalide F

Drug Discovery in Ocean. Figure 6
Kahalalide F, A depsipeptide produced by an algal endophytic marine bacterium

activity. In addition, Suarez et al. [81] reported that kahalalide F induced cell death via "oncosis" (defined as the progression of cellular processes leading to necrotic cell death) possibly initiated by lysosomal membrane depolarization in both prostate and breast cancer cell lines. This was followed in 2005 by a report by Sewell et al. [82] that HepG2 cells treated with 300 nM kahalalide F demonstrated significant alterations in their membrane permeability, which was manifested as cell swelling and/or blebbing, thus implying specific interactions with membranes and/or proteins at this concentration. In 2005, it was reported that it induced a necrosis-like cell death that involved inhibition of Akt signaling and depletion of ErbB3 [83]. Extension of this work was reported in 2006, when it was suggested that ErbB3 may well be a marker for progress against suitable tumor types in patients and also implied that an ErbB3 kinase inhibitor may well increase efficacy [52].

In 2005, a PCT International Application was filed by Hill et al. [84], claiming production of kahalalide F and other derivatives from a Vibrio species isolated from *Bryopsis* and also *Elysia rufesens*. This report implied that the invertebrate obtained the producing microbe from the alga and then maintained the microbe(s) as symbionts [84]. Accordingly, there may well be a potential renewable source of these agents by use of microbial fermentation.

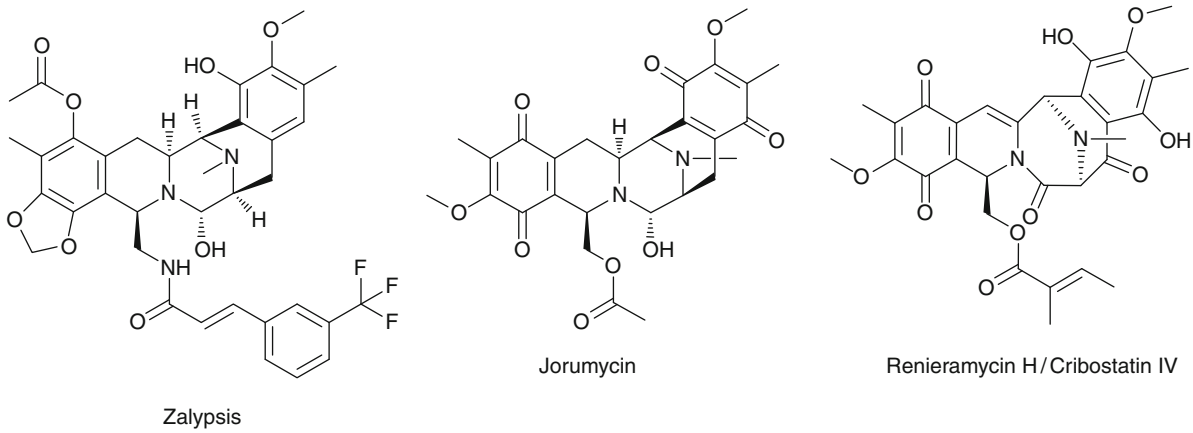
Although the compound went into phase II, clinical trials for cancer, it is currently in phase II clinical trials as a potential treatment for severe psoriasis, and over the last 2 years, two interesting papers have been published by PharmaMar scientists and/or their Spanish collaborators demonstrating that it is relatively easy to make over a 100 analogs with a variety of in vitro activities using solid state peptide synthesis techniques once the base structure was synthesized [85].

In addition to the antitumor and antipsoriasis activities reported above, it turns out that as well as having antifungal activity reported by one of the original discoverers of the class [86], in 2009, Cruz et al. reported that the molecule is also active against leishmanial infections [87].

Finally, Hawaii is not the only area with these peptides. Ashour et al. [88] reported new cyclic peptides of this class from the Indian Ocean mollusk, *Elysia grandifolia* and claimed a higher cytotoxicity in vitro than was reported for kahalalide F.

Zalypsis® (PM00104) This molecule (Fig. 7) is a "chemical cousin" of ET-743 (Yondelis®) though it is based upon the structures of two similar compound classes found in either mollusks (Jorumycin) (Fig. 7) or sponges (Renieramycin) (Fig. 7) [89–91]. PM00104 (Fig. 7) was derived from the base structure by subtle modifications and placed into preclinical and clinical trials by PharmaMar. It is currently in phase II trials against Ewing's sarcoma and endometrial and cervical cancers under PharmaMar auspices.

A fair number of papers dealing with the biochemical and pharmacological properties of PM00104 have recently been published [92, 93] together with an interesting report on the cross-resistance patterns demonstrated with paclitaxel and doxorubicin-resistant lines. It was shown that the expression of selected resistance markers does not correlate with the patterns shown



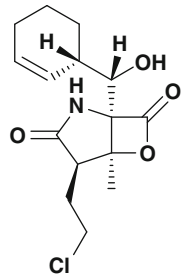
Drug Discovery in Ocean. Figure 7
Chemical cousins of trabectedin

with trabectedin and PM00104, and thus, subtly different genetic processes are involved with these two chemically similar compounds [94]. In addition, Pommier's group recently demonstrated that this compound and ET-743 differ in their interaction with DNA, with PM00104 demonstrating induction of a γ-H2AX response at nanomolar concentrations in Ewing's sarcoma cell lines; this observation suggests that PM00104 could serve as a pharmacodynamic marker for further clinical development of this agent in that particular cancer [95].

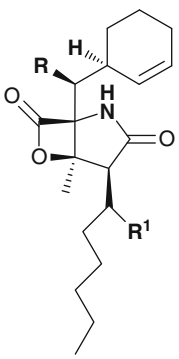
Salinosporamide A Salinosporamide A (Fig. 8) and the number of "chemical relatives" that have been reported since 2003 are vindication of an hypothesis that Fenical and Jensen had in the late 1980s to early 1990s that free-living microbes in the ocean or ocean sediments, in addition to those intimately involved with invertebrates, could produce compounds with pharmacologic potential. As in the case of the compounds descried earlier, due primarily to the source(s) of funding being either the NCI or related institutes at NIH, most of the investigations were directed towards antitumor agents from a practical aspect; however, the organisms were also studied from a microbial perspective by looking at the positions of isolated viable microbes in phylogenetic trees and, even more importantly, the basic genomic sequence(s) of metabolite-producing microbes from these pioneering studies as time progressed and techniques became available.

Salinosporamide A and its mechanism of action as a novel proteasome inhibitor was first reported by the Fenical and Jensen laboratories at the Scripps Institute of Oceanography in 2003 [96]; it was subsequently licensed to a small biotech company, Nereus Pharmaceuticals in San Diego, California for development. The compound had an unusual chlorine substitution pattern and within a year or so of the original publication, two academic groups had synthesized the base molecule and this was soon followed by a synthesis paper from Nereus scientists [97]. Subsequently, many groups reported improved syntheses with an excellent 2007 review covering almost all of the early studies [98].

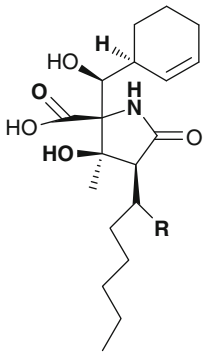
In what was a highly significant breakthrough, the cooperation between Nereus scientists and the then fermentation group at IRL in New Zealand led to the production of the necessary current Good Manufacturing Practices (cGMP) product for clinical trials; this cGMP involves large-scale fermentation in a saline environment, the first time that this task had been successfully performed on any scale with a marine microbe. During these preparative studies, a significant number of other salinosporamide derivatives and other secondary metabolites were further explored [99–103]. Concomitantly, the mechanism of salinosporamide as a proteasome inhibitor was worked out using excellent X-ray crystallographic studies [104] together with studies on the stability of the β-lactone ring [105].



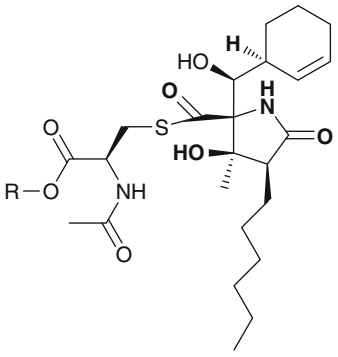
Salinosporamide A



Cinnabaramide A; R = OH, R^1 = H
Cinnabaramide B; R = R^1 = OH
Cinnabaramide C; R = R^1 = H



Cinnabaramide D; R = OH
Cinnabaramide E; R = H



Cinnabaramide F; R = H
Cinnabaramide G; R = CH_3

Drug Discovery in Ocean. Figure 8
Marine and terrestrial proteasome inhibitors from microbes

In 2007, an interesting paper [106] was published on the isolation of similar molecules, the cinnabaramides A-G (Fig. 8), from a terrestrial microbial source, which was followed the same year by a number of papers on the genomic aspects of the salinosporamide A producing organism, *Salinispora tropica*. Inspection of the complete genomic sequence of *S. tropica* demonstrated that these marine streptomycetes, as with their terrestrial cousins, have many more "currently unexpressed" secondary metabolite clusters in their genomes, thus demonstrating that they are as biochemically diverse as the quintessential terrestrial microbe *S. coelicolor* [107, 108]. It was further reported that the *S. tropica* genome encodes for a very unusual chlorinase that can be substituted by other halogens [109].

Within the last 2 years, three excellent reviews have been published on salinosporamides. Two cover the initial discovery and drug potential from the industrial (Nereus Pharma) aspect [110, 111] and the other covers biosynthetic and genomic processes as well as syntheses [112]. Both of the 2010 reviews should be read by those interested in this subject as they complement each other.

In a recent publication in the journal *Blood*, Nereus scientists and collaborators at the Harvard Medical School reported that salinosporamide demonstrated both in vitro and in vivo synergistic activities with lenalidomide (the thalidomide analog currently in clinical use) in multiple myeloma models, thus, potentially expanding clinical models with salinosporamide in due course [113].

Salinosporamide A entered phase I clinical trials in May of 2006 initially against solid tumors and leukemias and in April of 2007, another phase I trial against multiple myeloma was initiated. Currently, there are four phase I trials still listed as recruiting in the clinical trials data base, with one involving the use of the

histone deacetylase (HDAC) inhibitor vorinostat in conjunction with salinosporamide A.

Anti-infective Agents

Introduction

To date, there have been no anti-infective agents from marine sources that have been approved by any governmental organization equivalent to the US FDA. However, over the last 10 plus years, as funding became available, scientists have extended their searches for biologically active metabolites from marine sources into the areas of antiviral and antibacterial agents with most of the initial systematic studies being offshoots of antitumor investigations. Later studies are now being funded in anti-infective areas as a result of some of these earlier results.

In this section, antiviral agents and antiparasitic agents are dealt with, and the section finishes with studies on antimicrobial agents against both bacteria and fungi. It again needs to be emphasized that these are not yet clinical studies. In some cases, particularly in the case of anti-HIV agents, some materials are close to a version of clinical trials.

Antiviral Agents

Griffithsin This agent, first reported in 2005, was isolated from the red alga *Griffithsia* sp., and it was shown to be a 121-residue peptide that was subsequently produced by transfer of the DNA sequence corresponding to the peptide and expressed in *E. coli* [114]. The material, either natural or recombinant, was ultimately shown to bind to specific mannose-rich regions of HIV viral proteins [115]. In 2009, it was reported that the compound could function as a topical microbicide with potential as an intravaginal agent to protect against HIV transmission [116]. Since that publication, O'Keefe et al. [117] have continued their work with griffithsin and demonstrated that it has significant activity against a variety of human viral diseases, including the SARS virus [117]. In addition, recently, Moulaei et al. [118] further elucidated its activity against HIV entry. Then, from meeting reports that are not formally cited, griffithsin was reported to have activity against Ebola in a murine in vivo model [119].

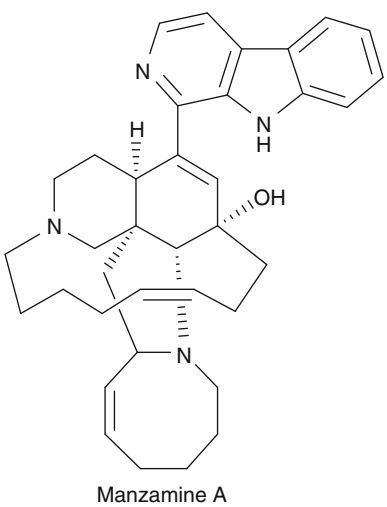
Antiparasitic Agents

Salinosporamide A In a paper published in 2008, Prudhomme et al., following testing of over 80 bacterial extracts and the pure marine-derived compound salinosporamide A (Fig. 8), demonstrated that this agent had significant activity against the erythrocytic stage of malaria, probably due to inhibition of the parasite's proteasome with an overall activity comparable to that of regular antimalarials such as chloroquin [120].

Other Agents In 2009, Fattoruso and Taglialatela-Scafati published an excellent review of the many marine-derived structures, including manzamine A (Fig. 9) derivatives that have demonstrated activity either in in vitro *or in vivo* models relevant to malaria [121]. Manzamine A (Fig. 9) and derivatives, though originally from a deepwater Indonesian sponge, are almost certainly produced by a commensal microbe and have shown activity in vivo against malarial models [122]. To date, however, no agents from marine sources have entered clinical trials.

Antibacterial or Fungal Agents

Since testing for antibacterial activities is a relatively simple and cheap operation, at least at the initial stage



Manzamine A

Drug Discovery in Ocean. Figure 9
Potential antiparasitic base structure from a deep water sponge

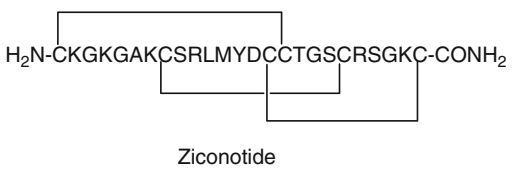
with crude extracts (simple disk diffusion assays against microbial lawns), a large number of groups have reported some antimicrobial activities for marine-derived structures. However, apart from some small companies and academic groups, no systematic investigations have yet been reported with marine-derived agents.

Fenical's group, among others, have realized the potential that is present in looking at invertebrate and marine microbial sourced extracts and compounds, and in a recent review article, they have demonstrated the potential for such a systematic investigation [123]. Obviously, they are not alone in this field. The 2007 review by Bull and Stach [124] and the very recent review on the biosynthetic potential of marine microbes from a genomic aspect must be read [125].

Ziconotide, Cone Snail Toxin

The fundamental work by Olivera et al. [126] on the peptide neurotoxins from fish-hunting cone snails led, over the next 20 years, to a massive amount of information as to the sources, utility in both the snail, and potential human use and to the first drug that was effectively directly from the sea as a human use pharmaceutical. This is the 25 residue highly cross-linked peptide known as ziconotide (Fig. 10), which was approved for severe chronic pain in 2005. Although ziconotide is made synthetically, it is chemically and biologically identical to the natural toxin.

The role of the *Conus* peptides in nature is as a method of paralyzing prey in order for the snail to be able to feed. From the information developed, mainly based on the pioneering work of Olivera's group, it is now only a matter of time before other cone snail-based drugs against a variety of disease of man will be coming out of clinical trials. That one can modify the base structure(s) and still maintain the desired activities, can be seen by the report from Craik's laboratories in Australia [127, 128], and when coupled to their predictive database of structures and potential pharmacology, the opportunities for the future development are significant [129].



Drug Discovery in Ocean. Figure 10
Cone snail toxin. Letters correspond to amino acids (www.biology.arizona.edu/biochemistry/problem_sets/aa/Dayhoff.html)

Future Directions

By its very nature, this article can only just "scratch the surface" of the potential for drug discovery from marine sources. As the understanding of the involvement of microbes of all domains expands, then it is highly probable that over the next decade or two, marine-sourced materials will be the building blocks, both literally and as leads to novel chemical structures, that will add to the armamentarium of drugs in areas that are in desperate need of novel agents.

If one looks at the multitude of reports on resistance to drugs by disease entities in anti-infective and antitumor areas to name but a few, then the absolute requirement for novel agents with new MOAs is quite apparent. That the marine area is a viable source for searching for such agents against a multitude of human diseases is now quite apparent as even though the costs and difficulties are very high, a significant number of the compounds that have been referred to in this chapter are completely different in their structures from anything yet found in terrestrial environments.

Bibliography

Primary Literature

1. Ramirez-Llodra E, Brandt A, Danovaro R, De Mol B, Escobar E, German CR, Levin LA, Martinez Arbizu P, Menot L, Buhl-Mortensen P, Narayanaswamy BE, Smith CR, Tittensor DP, Tyler PA, Vanreusel A, Vecchione M (2010) Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. Biogeosciences 7:2851–2899
2. Arrieta JM, Arnaud-Haond S, Duarte CM (2010) What lies underneath: conserving the oceans' genetic resources. Proc Natl Acad Sci USA 107:18318–18324
3. Gage JD, Tyler PA (1991) Deep-sea biology: a natural history of organisms at the deep-sea floor. Cambridge University Press, Cambridge
4. Hagiwara H, Numata M, Konishi K, Oka Y (1965) Synthesis of nereistoxin and related compounds I. Chem Pharm Bull 13:253–260

5. Ruggieri GD (1976) Drugs from the sea. Science 194:491–497
6. Suckling CJ (1991) Chemical approaches to the discovery of new drugs. Sci Prog 75:323–359
7. Newman DJ, Cragg GM, Snader KM (2000) The influence of natural products upon drug discovery. Nat Prod Rep 17: 215–234
8. Bergmann W, Feeney RJ (1950) Isolation of a new thymine pentoside from sponges. J Am Chem Soc 72:2809–2810
9. Bergmann W, Feeney RJ (1951) Marine products XXXII. The nucleosides of sponges. I. J Org Chem 16:981–987
10. Bergmann W, Burke DC (1955) Marine products XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. J Org Chem 20:1501–1507
11. Flahive E, Srirangam J (2005) The dolastatins: novel antitumor agents from *Dolabella auricularia*. In: Cragg GM, Kingston DGI, Newman DJ (eds) Anticancer agents from natural products. Taylor and Francis, Boca Raton, pp 191–213
12. Newman DJ (2005) The bryostatins. In: Cragg GM, Kingston DGI, Newman DJ (eds) Anticancer agents from natural products. Taylor and Francis, Boca Raton, pp 137–150
13. Hale KJ, Manaviazar S (2010) New approaches to the total synthesis of the bryostatin antitumor macrolides. Chem Asian J 5:704–754
14. Alkon DA, Sun M-K, Nelson TJ (2006) PKC signaling deficits: a mechanistic hypothesis for the origins of Alzheimer's disease. Trends Pharmcal Sci 28:51–60
15. Etcheberrigaray R, Tan M, Dewachtert I, Kuipéri C, Van Der Auwera I, Wera S, Qiao L, Bank B, Nelson TJ, Kozikowski AP, Van Leuven F, Alkon DL (2004) Therapeutic effects of PKC activators in Alzheimer's disease transgenic mice. Proc Natl Acad Sci USA 101:11141–11146
16. Montaner J (2006) Latest advances and research in stroke: focus on diagnostic and therapeutic targets. Drug News Perspect 19:173–193
17. Sun M-K, Alkon DL (2006) Bryostatin-1: pharmacology and therapeutic potential as a CNS drug. CNS Drug Rev 12:1–8
18. Wang D, Darwish DS, Schreurs BG, Alkon DL (2008) Analysis of long-term cognitive-enhancing effects of bryostatin-1 on the rabbit (*Oryctolagus cuniculus*) nictitating membrane response. Behav Pharmacol 19:245–256
19. Khan TK, Nelson TJ, Verma VA, Wender PA, Alkon DA (2009) A cellular model of Alzheimer's disease therapeutic efficacy: PKC activation reverses A beta-induced biomarker abnormality on cultured fibroblasts. Neurobiol Dis 34:332–339
20. Davidson SK, Allen SW, Lim GE, Anderson CM, Haygood MG (2001) Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus* Endobugula sertula" of the bryozoan *Bugula neritina*. Appl Environ Microbiol 67: 4531–4537
21. Davidson SK, Haygood MG (1999) Identification of sibling species of the bryozoan *Bugula neritina* that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont "*Candidatus* Endobugula sertula". Biol Bull 196:273–280
22. Haygood MG (2003) The role of a bacterial symbiont in the biosynthesis of bryostatins in the marine bryozoan *Bugula neritina*. Abstract paper 6th international marine biotechnology conference abstract S5-2A-K
23. Haygood MG, Davidson SK (1997) Small-subunit rRNA genes and in situ hybridization with oligonucleotides specific for the bacterial symbionts in the larvae of the bryozoan *Bugula neritina* and proposal of 'Candidatus Endobugula sertula'. Appl Environ Microbiol 63:4612–4616
24. Hildebran M, Waggoner LE, Lim GE, Sharp KH, Ridley CP, Haygood MG (2004) Approaches to identify, clone and express symbiont bioactive metabolite genes. Nat Prod Rep 21:122–142
25. Hildebrand M, Waggoner LE, Liu H, Sudek S, Allen S, Anderson C, Sherman DH, Haygood M (2004) *bryA*: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. Chem Biol 11:1543–1552
26. Lim GE, Haygood MG (2004) *Candidatus* Endobugula glebosa, a specific bacterial symbiont of the marine bryozoan *Bugula simplex*. Appl Environ Microbiol 70(8):4921–4929
27. Lim-Fong GE, Regali LA, Haygood MG (2008) Evolutionary relationships of "*Candidatus Endobugula*" bacterial symbionts and their *Bugula* bryozoan hosts. Appl Environ Microbiol 74:3605–3609
28. Lopanik N, Targett NM, Lindquist N (2006) Isolation of two polyketide synthase gene fragments from the unclutured microbial symbiont of the marine bryozoan *Bugula neritina*. Appl Environ Microbiol 72:7941–7944
29. Lopanik NB, Shields JA, Buchholz TJ, Rath CM, Hothersall J, Haygood MG, Håkansson K, Thomas CM, Sherman DH (2008) In vivo and in vitro trans-acylation by *bryP*, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. Chem Biol 15:1175–1186
30. Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, Patel A, Sherman DH, Haygood MG (2007) Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus Endobugula sertula*", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. J Nat Prod 70:67–74
31. Zimmermann K, Engeser M, Blunt JW, Munro MHG, Piel J (2009) Pederin-type pathways of uncultivated bacterial symbionts: analysis of O-methyltransferases and generation of a biosynthetic hybrid. J Am Chem Soc 131:2780–2781
32. Mitchell SS, Faulkner DJ, Rubins K, Bushman FD (2000) Dolastatin 3 and two novel cyclic peptides from a palauan collection of *Lyngbya majuscula*. J Nat Prod 63:279
33. Luesch H, Moore RE, Paul VJ, Mooberry SL, Corbett TH (2001) Isolation of dolastatin 10 from the marine cyanobacterium *Symploca* species VP642 and total stereochemistry and biological evaluation of its analogue symplostatin 1. J Nat Prod 64:907
34. Harrigan GG, Yoshida WY, Moore RE, Nagle DG, Park PU, Biggs J, Paul VJ, Mooberry SL, Corbett TH, Valeriote FA (1998) Isolation, structure determination, and biological activity of dolastatin 12

and lyngbyastatin I from *Lyngbya majuscula/Schizothrix calcicola* cyanobacterial assemblages. J Nat Prod 61:1221
35. Nogle LM, Williamson RT, Gerwick WH (2001) Somamides A and B, two new depsipeptide analogues of dolastatin 13 from a Fijian cyanobacterial assemblage of *Lyngbya majuscula* and *Schizothrix* species. J Nat Prod 64:716
36. Nogle LM, Gerwick WH (2002) Isolation of four new cyclic depsipeptides, antanapeptins A-D, and dolastatin 16 from a Madagascan collection of *Lyngbya majuscula*. J Nat Prod 65:21
37. Harrigan GG, Luesch H, Yoshida WY, Moore RE, Nagle DG, Paul VJ, Mooberry SL, Corbett TH, Valeriote FA (1998) Symplostatin 1: a dolastatin 10 analogue from the marine cyanobacterium *Symploca hydnoides*. J Nat Prod 61:1075
38. Harrigan GG, Luesch H, Yoshida WY, Moore RE, Nagle DG, Paul VJ (1999) Symplostatin 2: a dolastatin 13 analogue from the marine cyanobacterium *Symploca hydnoides*. J Nat Prod 62:655
39. Luesch H, Yoshida WY, Moore RE, Paul VJ, Mooberry SL, Corbett TH (2002) Symplostatin 3, a new dolastatin 10 analogue from the marine cyanobacterium *Symploca* sp VP452. J Nat Prod 65:16
40. Siemann DW (2011) The unique characteristics of tumor vasculature and preclinical evidence for its selective disruption by Tumor-Vascular Disrupting Agents. Cancer Treat Rev 31:63–74
41. Patel S, Keohan ML, Saif MW, Rushing D, Baez L, Feit K, DeJager R, Anderson S (2006) Phase II study of intravenous TZT-1027 in patients with advanced or metastatic soft-tissue sarcomas with prior exposure to anthracycline-based chemotherapy. Cancer 107:2881–2887
42. Riely GJ, Gadgeel S, Rothman I, Saidman B, Sabbath K, Feit K, Kris MG, Rizvi NA (2007) A phase 2 study of TZT-1027, administered weekly to patients with advanced non-small cell lung cancer following treatment with platinum-based chemotherapy. Lung Cancer 55(2):181–185
43. Carter NJ, Keam SJ (2010) Trabectedin: a review of its use in soft tissue sarcoma and ovarian cancer. Drugs 70(3):355–376
44. Poveda A, Vergote I, Tjulandin S, Kong B, Roy M, Chan S, Filipczyk-Cisarz E, Hagberg H, Kaye SB, Colombo N, Lebedinsky C, Parekh T, Gómez J, Park YC, Alfaro V, Monk BJ (2011) Trabectedin plus pegylated liposomal doxorubicin in relapsed ovarian cancer: outcomes in the partially platinum-sensitive (platinum-free interval 6–12 months) subpopulation of OVA-301 phase III randomized trial. Ann Oncol 22:39–48
45. Rinehart K, Holt TG, Fregeau NL, Stroh JG, Kiefer PA, Sun F, Li LH, Martin DG (1990) Ecteinascidins 729, 743, 745, 759A, 759B and 770: potent antitumor agents from the Caribbean tunicate *Ecteinascidia turbinata*. J Org Chem 55:4512–4515
46. Wright AE, Forleo DA, Gunawardana GP, Gunasekera SP, Koehn FE, McConnell OJ (1990) Antitumor tetrahydroisoquinoline alkaloids from the colonial ascidian *Ecteinascidia turbinata*. J Org Chem 55:4508–4512
47. Henriquez R, Faircloth G, Cuevas C (2005) Ecteinascidin 743 (ET-743; Yondelis™), aplidin, and kahalalide F. In: Cragg GM, Kingston DGI, Newman DJ (eds) Anticancer agents from natural products. Taylor and Francis, Boca Raton, pp 215–240
48. Velasco A, Acebo P, Gomez A, Schleissner C, Rodriguez P, Aparicio T, Conde S, Munoz R, de la Calle F, Garcia JL, Sanchez-Puelles JM (2005) Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytoxic compounds. Mol Microbiol 56:144–154
49. Moss C, Green DH, Perez B, Velasco A, Henriquez R, McKenzie JD (2003) Intracellular bacteria associated with the ascidian *Ecteinascidia turbinata*: phylogenic and in situ hybridization analysis. Mar Biol 143:99–110
50. Perez-Matos AE, Rosado W, Govind NS (2007) Bacterial diversity associated with the Caribbean tunicate *Ecteinascidia turbinata*. Antonie van Leeuwenhoek 92:155–164
51. D'Incalci M, Galmarini CM (2010) A review of trabectedin (ET-743): a unique mechanism of action. Mol Cancer Ther 9:2157–2163
52. Jimeno J, Aracil M, Tercero JC (2006) Adding pharmacogenomics to the development of new marine-derived anticancer agents. J Transl Med 4:3
53. Soares DG, Escargueil AE, Poindessous V, Sarasin A, de Gramont A, Bonatto D, Henriques JAP, Larsen AK (2007) Replication and homologous recombination repair regulate DNA double-strand break formation by the antitumor alkylator ecteinascidin 743. Proc Natl Acad Sci USA 104:13062–13067
54. Marco E, David-Cordonnier M-H, Bailly C, Cuevas C, Gago F (2006) Further insight into the DNA recognition mechanism of trabectedin from the differential affinity of its demethylated analogue ecteinascidin ET729 for the triplet DNA binding site CGA. J Med Chem 49:6925–6929
55. Moneo V, Serelde BG, Fominaya J, Leal JFM, Blanco-Aparicio C, Romero L, Sanchez-Beato M, Cigudosa JC, Tercero JC, Piris MA, Jimeno J, Carnero A (2007) Extreme sensitivity to Yondelis(R) (Trabectedin, ET-743) in low passaged sarcoma cell lines correlates with mutated p53. J Cell Biochem 100:339–348
56. Aicher TD, Buszek KR, Fang FG, Forsyth CJ, Jung SH, Kishi Y, Matelich MC, Scola PM, Spero DM, Yoon SK (1992) Total synthesis of halichondrin B and norhalichondrin B. J Am Chem Soc 114:3162–3164
57. Yu MJ, Kishi Y, Littlefield BA (2005) Discovery of E7389, a fully synthetic macrocyclic ketone analog of halichondrin B. In: Cragg GM, Kingston DGI, Newman DJ (eds) Anticancer agents from natural products. Taylor and Francis, Boca Raton, pp 241–265
58. Jordan MA, Kamath K, Manna T, Okouneva T, Miller HP, Davis C, Littlefield BA, Wilson L (2005) The primary antimitotic mechanism of action of the synthetic halichondrin E7389 is suppression of microtubule growth. Mol Cancer Ther 4:1086–1095
59. Jordan MA, Kamath K (2007) How do microtubule-targeted drugs work? An overview. Curr Cancer Drug Targets 7:730–742
60. Okouneva T, Azarenko O, Wilson L, Littlefield BA, Jordan MA (2008) Inhibition of centromere dynamics by Eribulin (E7389) during mitotic metaphase. Mol Cancer Ther 7:2003–2011

61. Dabydeen DA, Burnett JC, Bai R, Verdier-Pinard P, Hickford SJH, Pettit GR, Blunt JW, Munro MHG, Gussio R, Hamel E (2006) Comparison of the activities of the truncated halichondrin B analog NSC 707389 (E7389) with those of the parent compound and a proposed binding site on tubulin. Mol Pharmacol 70:1866–1875
62. Allday PH, Correia JJ (2009) Macromolecular interaction of halichondrin B analogues eribulin (E7389) and ER-076349 with tubulin by analytical ultracentrifugation. Biochemistry 48:7927–7938
63. Smith JA, Wilson L, Azarenko O, Zhu X, Lewis BM, Littlefield BA, Jordan MA (2010) Eribulin binds at microtubule ends to a single site on tubulin to suppress dynamic instability. Biochemistry 49:1331–1337
64. Dong C-G, Henderson JA, Kaburagi Y, Sasaki T, Kim D-S, Kim JT, Urabe D, Guo H, Kishi Y (2009) New syntheses of E7389 C14-C35 and halichondrin C14-C38 building blocks: reductive cyclization and oxy-Michael cyclization approaches. J Am Chem Soc 131:15642–15646
65. Kim D-S, Dong C-G, Kim JT, Guo H, Huang J, Tiseni PS, Kishi Y (2009) New syntheses of E7389 C14-C35 and halichondrin C14-C38 building blocks: double-Inversion approach. J Am Chem Soc 131:15636–15641
66. Yang Y-R, Kim D-S, Kishi Y (2009) Second generation synthesis of C27-C35 building block of E7389, a synthetic halichondrin analogue. Org Lett 11:4516–4519
67. Wang Y, Serradell N, Bolos J, Rosa E (2007) Eribulin mesilate. Drugs Future 32:681–698
68. Jackson KL, Henderson JA, Phillips AJ (2009) The halichondrins and E7389. Chem Rev 109:3044–3079
69. Jackson KL, Henderson JA, Motoyoshi H, Phillips AJ (2009) A total synthesis of norhalichondrin B. Angew Chem Int Ed 48:2346–2350
70. Narayan S, Carlson EM, Cheng H, Du H, Hu Y, Jiang Y, Lewis BM, Seletsky BM, Tendyke K, Zhang H, Zheng W, Littlefield BA, Towle MJ, Yu MJ (2011) Novel second generation analogs of eribulin. Part I: compounds containing a lipophilic C32 sidechain overcome P-glycoprotein susceptibility. Bioorg Med Chem Lett 21:1630–1633
71. Narayan S, Carlson EM, Cheng H, Condon K, Du H, Eckley S, Hu Y, Jiang Y, Kumar V, Lewis BM, Saxton P, Schuck E, Seletsky BM, Tendyke K, Zhang H, Zheng W, Littlefield BA, Towle MJ, Yu MJ (2011) Novel second generation analogs of eribulin. Part II: orally available and active against resistant tumors in vivo. Bioorg Med Chem Lett 21:1634–1638
72. Narayan S, Carlson EM, Cheng H, Condon K, Du H, Eckley S, Hu Y, Jiang Y, Kumar V, Lewis BM, Saxton P, Schuck E, Seletsky BM, Tendyke K, Zhang H, Zheng W, Littlefield BA, Towle MJ, Yu MJ (2011) Novel second generation analogs of eribulin. Part III: blood-brain barrier permeability and in vivo activity in a brain tumor model. Bioorg Med Chem Lett 21:1639–1643
73. Sakai R, Rinehart KL, Kishore V, Kundu B, Faircloth G, Gloer JB, Carney JR, Manikoshi M, Sun F, Hughes RG Jr, Garcia-Gravalos D, Garcia de Quesada T, Wilson GR, Heid RM (1996) Structure-activity relationships of the didemnins. J Med Chem 39:2819–2834
74. Broggini M, Marchini SV, Galliera E, Borsotti P, Taraboletti G, Erba E, Sironi M, Jimeno J, Faircloth GT, Giavazzi R, D'Incalci M (2003) Aplidine, a new anticancer agent of marine origin, inhibits vascular endothelial growth factor (VEGF) secretion and blocks VEGF-VEGFR-1 (*flt-1*) autocrine loop in human leukemia cells MOLT-4. Leukemia 17:52–59
75. Erba E, Serafini M, Gaipa G, Tognon G, Marchini S, Celli N, Rotilio D, Broggini M, Jimeno J, Faircloth GT, Biondi A, D'Incalci M (2003) Effect of Aplidin in acute lymphoblastic leukaemia cells. Br J Cancer 89:763–773
76. Munoz-Alonzo MJ, Gonzalez-Santiago L, Martinez T, Losada A, Galmarini CM, Munoz A (2009) The mechanism of action of plitidepsin. Curr Opin Investig Drugs 10:536–542
77. Vera M, Joullie MM (2002) Natural products as probes of cell biology: 20 years of didemnin research. Med Res Rev 22:102–145
78. Hamann MT, Otto CS, Scheuer PJ, Dunbar DC (1996) Kahalalides: bioactive peptides from a marine mollusk *Elysia rufescens* and its algal diet *Bryopsis* sp. J Org Chem 61:6594–6600
79. Lopez-Macia A, Jimenez JC, Royo M, Giralt E, Alberico F (2001) Synthesis and structure determination of kahalalide F. J Am Chem Soc 123:11398–11401
80. Garcia-Rocha M, Bonay P, Avila J (1996) The antitumoral compound kahalalide F acts on cell lysosomes. Cancer Lett 99:43–50
81. Suarez Y, Gonzalez L, Cuadrado A, Berciano M, Lafarga M, Munoz A (2003) Kahalalide F, a new marine-derived compound, induces oncosis in human prostate and breast cancer cells. Mol Cancer Ther 2:863–872
82. Sewell JM, Mayer I, Langdon SP, Smyth JF, Jodrell DI, Guichard SM (2005) The mechanism of action of kahalalide F: variable cell permeability in human hepatoma cell lines. Eur J Cancer 41:1637–1644
83. Janmaat ML, Rodriguez JA, Jimeno J, Kruyt FAE, Giaccone G (2005) Kahalalide F induces necrosis-like cell death that involves depletion of ErbB3 and inhibition of Akt signaling. Mol Pharmacol 68:502–510
84. Hil RT, Hamann MT, Enticknap J, Karumanchi V (2005) Kahalalide-producing bacteria and methods of identifying kahalalide-producing bacteria and producing kahalalides. WO 2005042720
85. Jimenez JC, Lopez-Macia A, Gracia C, Varon S, Carrascal M, Caba JM, Royo M, Francesch AM, Cuevas C, Giralt E, Alberico F (2008) Structure-activity relationship of kahalalide F synthetic analogues. J Med Chem 51:4920–4931
86. Shilabin AG, Kasanah N, Wedge DE, Hamann MT (2007) Lysosome and HER3 (ErbB3) selective anticancer agent kahalalide F: semisynthetic modifications and antifungal lead exploration studies. J Med Chem 50:4340–4350
87. Cruz LJ, Luque-Ortega JR, Rivas L, Alberico F (2009) Kahalalide F, an antitumor depsipeptide in clinical trials, and its analogues as effective antileishmanial agents. Mol Pharm 6:813–824

88. Ashour M, Edrada RA, Ebel R, Wray V, Waetjen W, Padmakumar K, Mueller WEG, Lin WH, Proksch P (2006) Kahalalide derivatives from the Indian sacoglossan mollusk *Elysia grandifolia*. J Nat Prod 69:1547–1553
89. Fontana A, Cavaliere P, Wahidulla S, Naik CG, Cimino G (2000) A new antitumor isoquinoline alkaloid from the marine nudibranch *Jorunna funebris*. Tetrahedron 56:7305–7308
90. Saito N, Tanaka C, Koizumi Y-I, Suwanborirux K, Amnuoypol S, Pummangura S, Kubo A (2004) Chemistry of renieramycins. Part 6: transformation of renieramycin M into jorumycin and renieramycin J including oxidative degradation products, mimosamycin, renierone, and renierol acetate. Tetrahedron 60:3873–3881
91. Lane JW, Chen Y, William RM (2005) Asymmetric total syntheses of (−)-jorumycin, (−)-renieramycin G, 3-epi-jorumycin, and 3-epi-renieramycin G. J Am Chem Soc 127:12684–12690
92. Leal JFM, García-Hernández V, Moneo V, Domingo A, Bueren-Calabuig JA, Negri A, Gago F, Guillén-Navarro MJ, Avilés P, Cuevas C, García-Fernández LF, Galmarini CM (2009) Molecular pharmacology and antitumor activity of Zalypsis® in several human cancer cell lines. Biochem Pharmacol 78:162–170
93. Ocio EM, Maiso P, Chen X, Garayoa M, Álvarez-Fernández S, San-Segundo L, Vilanova D, López-Corral L, Montero JC, Hernández-Iglesias T, de Álava E, Galmarini C, Avilés P, Cuevas C, San-Miguel JF, Pandiella A (2009) Zalypsis: a novel marine-derived compound with potent antimyeloma activity that reveals high sensitivity of malignant plasma cells to DNA double-strand breaks. Blood 113:3781–3791
94. Duan Z, Choy E, Jimeno JM, Cuevas CD-M, Mankin HJ, Hornicek FJ (2009) Diverse cross-resistance phenotype to ET-743 and PM00104 in multi-drug resistant cell lines. Cancer Chemother Pharmacol 63:1121–1129
95. Guirouilh-Barbat J, Antony S, Pommier Y (2009) Zalypsis (PM00104) is a potent inducer of gamma-H2AX foci and reveals the importance of the C ring of trabectedin for transcription-coupled repair inhibition. Mol Cancer Ther 8:2007–2014
96. Feling RH, Buchanan GO, Mincer TJ, Kauffman CA, Jensen PR, Fenical W (2003) Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. Angew Chem Int Ed 42:355–357
97. Macherla VR, Mitchell SS, Manam RR, Reed KA, Chao T-H, Nicholson B, Deyanat-Yazdi G, Mai B, Jensen PR, Fenical WF, Neuteboom STC, Lam KS, Palladino MA, Potts BCM (2005) Structure-activity relationship studies of salinosporamide A (NPI-0052), a novel marine derived proteasome inhibitor. J Med Chem 48:3684–3687
98. Shibasaki M, Kanai M, Fukuda N (2007) Total synthesis of lactacystin and salinosporamide A. Chem Asian J 2:20–38
99. Jensen PR, Williams PG, Oh D-C, Zeigler L, Fenical W (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. Appl Environ Microbiol 73:1146–1152
100. Lam KS, Tsueng G, McArthur KA, Mitchell SS, Potts BCM, Xu J (2007) Effects of halogens on the production of salinosporamides by the obligate marine actinomycete *Salinispora tropica*. J Antibiot 60:13–19
101. Reed KA, Manam RR, Mitchell SS, Xu J, Teisan S, Chao T-H, Deyanat-Yazdi G, Neuteboom STC, Lam KS, Potts BCM (2007) Salinosporamides D-J from the marine actinomycete salinispora tropica, bromosalinosporamide, and thioester derivatives are potent inhibitors of the 20S proteasome. J Nat Prod 70:269–276
102. Tsueng G, Lam KS (2007) Stabilization effect of resin on the production of potent proteasome inhibitor NPI-0052 during submerged fermentation of *Salinispora tropica*. J Antibiot 60:469–472
103. Tsueng G, McArthur KA, Potts BCM, Lam KS (2007) Unique butyric acid incorporation patterns for salinosporamides A and B reveal distinct biosynthetic origins. Appl Microbiol Biotechnol 75:999–1005
104. Groll M, Huber R, Potts BCM (2006) Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of beta-lactone ring opening and a mechanism for irreversible binding. J Am Chem Soc 128:5136–5141
105. Denora N, Potts BCM, Stella VJ (2007) A mechanistic and kinetic study of the β-lactone hydrolysis of salinosporamide A (NPI-0052), a novel proteasome inhibitor. J Pharm Sci 96:2037–2047
106. Stadler M, Bitzer J, Mayer-Bartschmid A, Mueller H, Benet-Buchholz J, Gantner F, Tichy H-V, Reinemer P, Bacon KB (2007) Cinnabaramides A-G: analogues of lactacystin and salinosporamide from a terrestrial streptomycete. J Nat Prod 70:246–252
107. Beer LL, Moore BS (2007) Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica*. Org Lett 9:845–848
108. Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proc Natl Acad Sci USA 104:10376–10381
109. Eustaquio AS, Pojer F, Noel JP, Moore BS (2008) Discovery and characterization of a marine bacterial SAM-dependent chlorinase. Nat Chem Biol 4:69–74
110. Fenical W, Jensen PR, Palladino MA, Lam KS, Lloyd GK, Potts BC (2009) Discovery and development of the anticancer agent salinosporamide A (NPI-0052). Bioorg Med Chem 17(6):2175–2180
111. Potts BC, Lam KS (2010) Generating a generation of proteasome inhibitors: from microbial fermentation to total synthesis of salinosporamide A (Marizomib) and other salinosporamides. Mar Drugs 8:835–880
112. Gulder TAM, Moore BS (2010) Salinosporamide natural products: potent 20S proteasome inhibitors as promising cancer chemotherapeutics. Angew Chem Int Ed 49: 9346–9367

113. Chauhan D, Singh AV, Ciccarelli B, Richardson PG, Palladino MA, Anderson KC (2010) Combination of novel proteasome inhibitor NPI-0052 and lenalidomide trigger in vitro and in vivo synergistic cytotoxicity in multiple myeloma. Blood 115:834–845
114. Mori T, O'Keefe BR, Sowder RC II, Bringans S, Gardella R, Berg S, Cochran P, Turpin JA, Buckheit RW Jr, McMahon JB, Boyd MR (2005) Isolation and characterization of Griffithsin, a novel HIV-inactivating protein, from the red alga *Griffithsia* sp. J Biol Chem 280:9345–9353
115. Ziółkowska NE, Shenoy SR, O'Keefe BR, Wlodawer A (2007) Crystallographic studies of the complexes of antiviral protein griffithsin with glucose and N-acetylglucosamine. Protein Sci 16:1485–1489
116. O'Keefe BR, Vojdani F, Buffa V, Shattock RJ, Montefiori DC, Bakke J, Mirsalis J, d'Andrea A-L, Hume SD, Bratcher B, Saucedo CJ, McMahon JB, Pogue GP, Palmer KE (2009) Scaleable manufacture of HIV-1 entry inhibitor griffithsin and validation of its safety and efficacy as a topical microbicide component. Proc Natl Acad Sci USA 106:6099–6104
117. O'Keefe BR, Giomarelli B, Barnard DL, Shenoy SR, Chan PKS, McMahon JB, Palmer KE, Barnett BW, Meyerholz DK, Wohlford-Lenane CL, McCray PB Jr (2010) Broad-spectrum in vitro activity and in vivo efficacy of the antiviral protein griffithsin against emerging viruses of the family coronaviridae. J Virol 84:2511–2521
118. Moulaei T, Shenoy SR, Giomarelli B, Thomas C, McMahon JB, Dauter Z, O'Keefe BR, Wlodawer A (2010) Monomerization of viral entry inhibitor griffithsin elucidates the relationship between multivalent binding to carbohydrates and anti-HIV activity. Structure 18:1104–1115
119. Service RF (2009) Sugary achilles heel raises hope for broad-acting antiviral drugs. Science 325:1200
120. Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, Jensen PR, Le Roch K (2008) Marine actinomycetes: a new source of compounds against the human malaria parasite. PLoS One 3:e2335
121. Fattoruso E, Taglialatela-Scafati O (2009) Marina antimalarials. Mar Drugs 7:130–152
122. Ang KKH, Holmes MJ, Higa T, Hamann MT, Kara UAK (2000) In vivo antimalarial activity of the beta-carboline alkaloid, manzamine A. Antimicrob Agents Chemother 44: 1645–1649
123. Hughes CC, Fenical W (2010) Antibacterials from the sea. Chem Eur J 16:12512–12525
124. Bull AT, Stach JEM (2007) Marine actinobacteria: new opportunities for natural product search and discovery. Trends Microbiol 15:491–499
125. Lane AL, Moore BS (2011) A sea of biosynthesis: marine natural products meet the molecular age. Nat Prod Rep 28:411–428
126. Olivera BM, Gray WR, Zeikus R (1985) Peptide neurotoxins from fish-hunting cone snails. Science 230:1338–1343
127. Craik DJ, Adams DJ (2007) Chemical modification of conotoxins to improve stability and activity. ACS Chem Biol 2:457–468
128. Hala R, Craik DJ (2009) Conotoxins: natural product drug leads. Nat Prod Rep 26:526–536
129. Kaas Q, Westermann J-C, Craik DJ (2010) Conopeptide characterization and classifications: an analysis using ConoServer. Toxicon 55:1491–1509

Books and Reviews

Abraham DJ, Rotella DP (2010) Burger's medicinal chemistry, drug discovery and development, 7th edn. Wiley, Hoboken
Buss AD, Butler MS (2010) Natural product chemistry for drug discovery. RSC Publishing, Cambridge, UK
Chivian E, Bernstein A (2008) Sustaining life: how human health depends on biodiversity. Oxford University Press, Oxford
Rydzewski RM (2008) Real world drug discovery: a chemist's guide to biotech and pharmaceutical research. Elsevier, Amsterdam

Dynamic Environment Sensing Using an Intelligent Vehicle

Huijing Zhao, Long Xiong, Yiming Liu, Xiaolong Zhu, Yipu Zhao, Hongbin Zha
Key Lab of Machine Perception (MOE), Peking University, Beijing, China

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Sensor and System Architecture
Sensor Calibration and Data Alignment
Contextual Map Generation and Representation
Future Directions
Bibliography

Glossary

Classification Annotating a data segment or an object by a class label.

Contextual map A map containing high-level knowledge beyond geometry, such as object, class, and motion.

Data alignment Integrating the instantaneous measurements from sensors' coordinate system(s) to a global coordinate system.

Dynamic mapping A mapping technology that uses a dynamic procedure or generates a map of a dynamic environment.

Multi-laser sensing system A sensor system that makes collaborative use of a number of laser range scanners.

Range image A 2D image where the value of each pixel is a range distance, with the pixel index corresponding to range angle and scanning sequence, so that a 3D coordinate can be retrieved for each pixel of the range image.

Scene understanding Converting from low-level knowledge to high-level knowledge of an environment.

Segmentation Making partitions on a data set, where in each partition cell (i.e., segment), data has the property of certain homogeneity.

Sensor calibration Finding a set of parameters that describes internal or external sensor geometry.

Definition of the Subject and Its Importance

There have been numerous research efforts toward generating 2D/3D urban models using mobile robots. In addition, research has focused on robot-centric mapping and moving object detection/tracking for online perception and navigation. However, to date, there has been little work on generating a realistic 3D copy of a dynamic environment that describes the state of both static and dynamic objects at the moment. Toward the goal of developing omni-directional range sensing in a dynamic urban scene using an intelligent vehicle, research has mainly focused on the fundamental issues of multi-laser sensor system calibration and scene understanding in contextual map generation. Here, both system and algorithmic development are presented as well as experimental research demonstrating that a geometric and contextual representation of static objects such as buildings, trees, and roads, as well as the motion of dynamic entities such as people, bicycles, and cars, can be achieved using a vehicle robot.

Introduction

Traditionally, urban modeling has been thought of as generating maps of cities, in either 2D or 3D, for applications such as city planning, facility management, tourism, navigation, and cultural heritage. Numerous research efforts in photogrammetry and remote sensing [1] have been devoted to the study of aerial or satellite-based mapping technologies for the reconstruction of urban objects [2–4], yielding a vast amount of geographic databases. A summary of worldwide urban modeling projects can be found in [5]. Normally, aerial sensing can cover relatively wide area, but it fails to capture urban details due to the limitation of spatial resolution and a bird's eye viewpoint. Mobile mapping systems (MMS), which emerged at the end of the last decade [6–8], were initially considered as a complimentary mapping approach to aerial ones. This technological trend is described in [9, 10]. In such systems, ground vehicles are used as moving platforms carrying multi-modal sensors, so that city objects can be measured from nearby viewpoints on street. Early results [11–13] demonstrate thatrich data sets can be acquired, and highly realistic city models can be reconstructed using such a mapping technique. With the success of Google Earth [14] and Microsoft Virtual Earth [15], a broader community has found interest in visualizing cities in multilevels of detail by combining data from both aerial and ground-based mapping techniques. Applications linking geographical maps with embedded 3D/2D visualization from street viewpoints are increasing. One of the most successful applications is Google StreetView [16], and a number of commercial systems using mobile mapping technologies [17–19] provide 3D/2D urban data. A review of the major approaches to large-scale modeling is given in [20]. However, these systems represent only the static entities of the environments, while the dynamic objects, such as cars and people, are either removed from the models or left on 2D images that do not reflect the current situation of the environment.

Mapping has been an active area of research in robotics and artificial intelligence for decades. Differing from photogrammetry and remote sensing, robotic mapping generates and maintains a spatial model of the robot's working environment, which is used to help

the robot's online perception and decision making. Maps vary in scale, dimension, content, and representation, according to the robot's function. A comprehensive survey of robotic mapping techniques, i.e. the key issues, major approaches and future challenges, is given in [21]. In recent years, interest in ground vehicle robots has grown rapidly. Especially with the DARPA Grand Challenges 2004, 2005 [22] and Urban Challenge 2007 [23], considerable enthusiasm and research interest have been inspired in the robotics and intelligent vehicle (IV) communities toward developing autonomous driving techniques and advanced driver assistance systems. The ground vehicle robot or intelligent vehicle must function in a complex outdoor environment [24, 25], where a map that eprovides knowledge of the highly dynamic environment is essential for the robot to achieve advanced online perception and decision. Generating maps of dynamic outdoor environments that contain the representations of both static and dynamic entities using mobile robots is still a fundamental issue.

There have been numerous research efforts to generate 2D/3D urban models using mobile robots [26–29]. There has also been work on robot-centric mapping and moving object detection/tracking for online perception and navigation [30–34]. However, to date, there has been little research on generating realistic 3D copies of a dynamic environment that describes the current situation of both static and dynamic objects at the moment. This research focuses on developing an intelligent vehicle system that conducts omnidirectional range sensing to a dynamic urban scene, where a geometric and contextual representation of static objects such as buildings, trees, and roads,as well as the motion at the moment of dynamic entities such as people, bicycles, and cars, can be achieved after a run. In the following, sensor and system architecture is first addressed, followed by the description to the two major modules, sensor and data alignment, and map generation. Algorithms and experimental results are presented, and finally future directions are discussed.

Sensor and System Architecture

Sensor System

Over the past decade, laser range scanners, either 2D or 3D, have emerged as the dominant exteroceptive sensor for mobile robotics applications. Instead of using the 3D laser scanners, such as Velogyne HDL-64E [35], which has 64 laser heads and can achieve omnidirectional 3D range sensing with a scan rate of 5–15 Hz, this research developed a multi-laser sensor system by integrating a number of single-row laser scanners (called LIDAR, for Light Detection and Ranging) in order to achieve low sensor cost and flexibility for a variety of applications. It is desirable that the software be adaptable to different numbers and layouts of laser scanners according to application requirements.

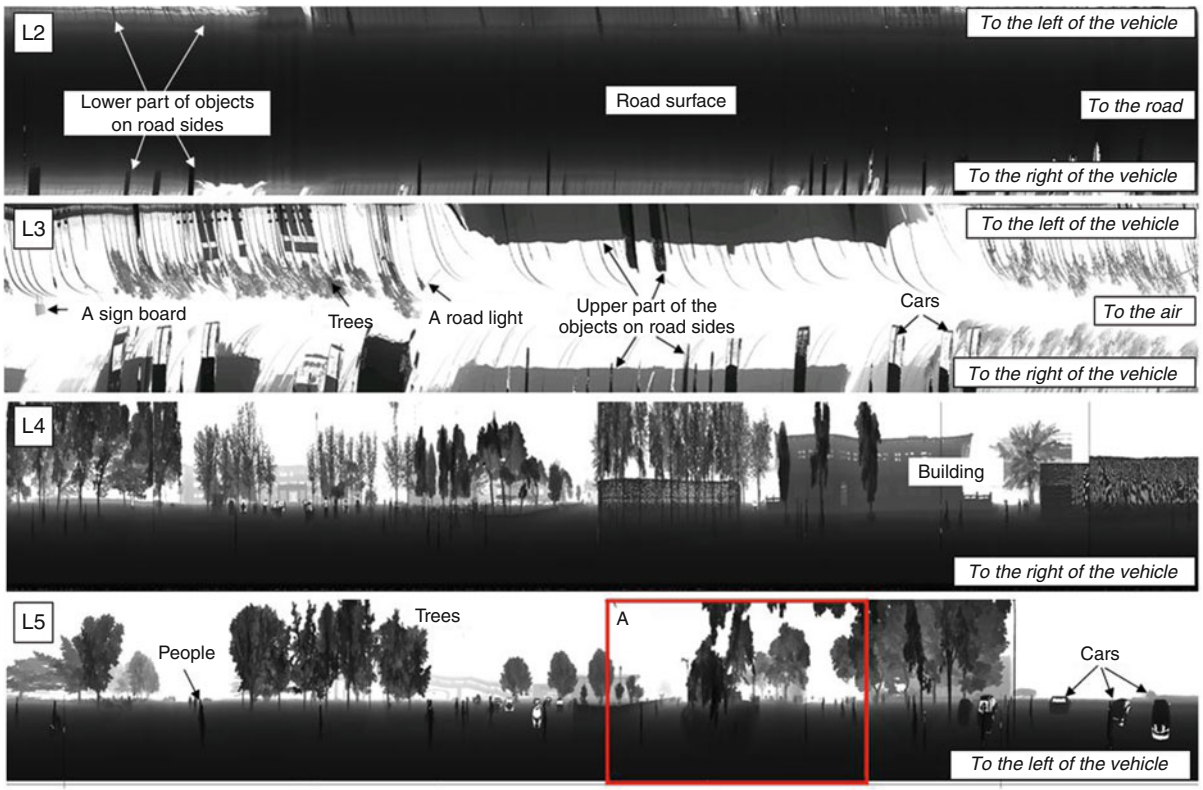
Figure 1 shows an example of sensor layout with five SICK LMS2**s [36] and a GPS (Global Positioning System)/IMU (Inertial Measurement Unit) based navigation unit. L1 is a horizontal scanning unit mounted in the center of the front bumper, which is used in combination with GPS/IMU so as to estimate vehicle position while generating a horizontal reference, i.e., a 2D grid map of the environment, and extracting motion trajectories of dynamic objects at a horizontal level (SLAMMODT [31]: Simultaneous Localization and Mapping [37, 38] with Moving Object Detection and Tracking [39, 40]). Other laser scanners are used to acquire 3D data about the environment along streets: L2 scans downwardly, measuring the lower part of objects on roadside, as well as road surface; L3 scans the upper part of objects on roadside, as well as those above the road surface, such as traffic signs and signal lights; L4 and L5 scan vertically to the left and right sides of the car. The data from each laser scanner can be represented in the form of 2D range images, as shown in Fig. 2, where the horizontal axis is time, and vertical strips correspond to laser scan lines, with each pixel valued by converting range distance into an intensity in [0, 255]. A piece of range segment, i.e., area "A" in Fig. 2, is enlarged in Fig. 3. It can be seen that a dynamic environment containing not only the stationary objects such as trees and poles, but also the moving objects such as pedestrians and cyclists, is vividly captured through such a range measurement.

Software Modules

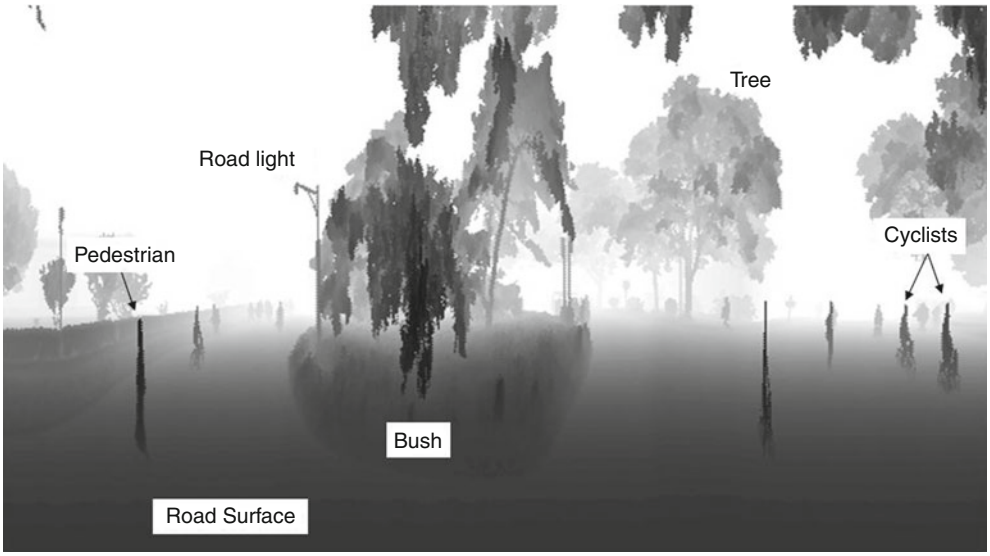
Figure 4 describes software modules and major data flows (input and output). Module 1 is a sensor data logger, which records both GPS/IMU navigation inputs

Laser #	Scan mode	Major purpose
L1	Horizontal	SLAM with MODT
L2	Downward	Road geometry
L3	Upward	Objects above the road
L4	Vertical Right	Objects to the right of the car
L5	Vertical Left	Objects to the left of the car

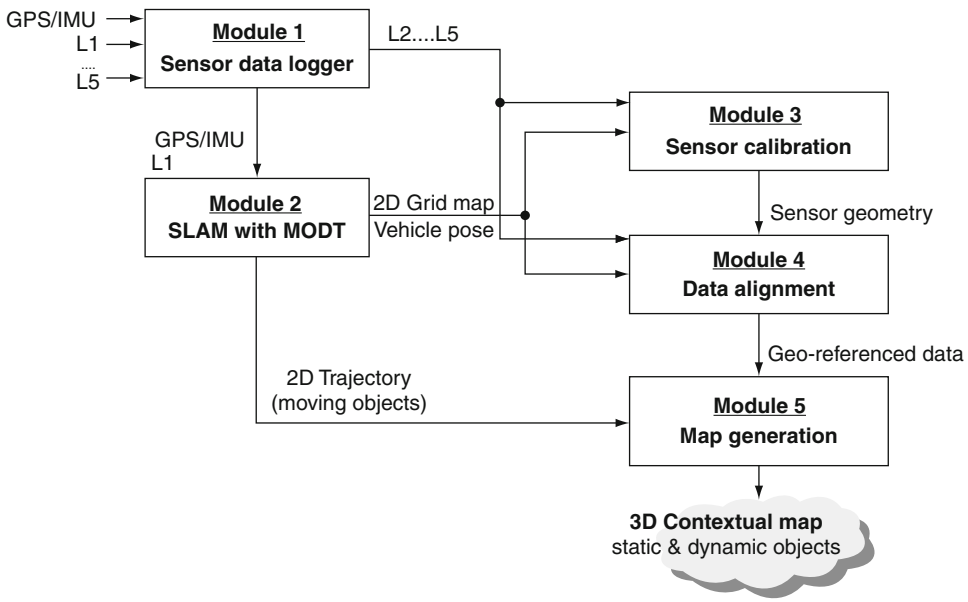
Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 1
A test-bed vehicle system designed for omni-directional range sensing [41]



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 2
Range images of laser scan data L2–L5 [42]. Horizontal axis is scan line number (corresponding to time). Each vertical strip represents a line of range values of a certain laser scan



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 3
An enlarged range image by laser scanner L5 [42]



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 4
Software modules and major data flows

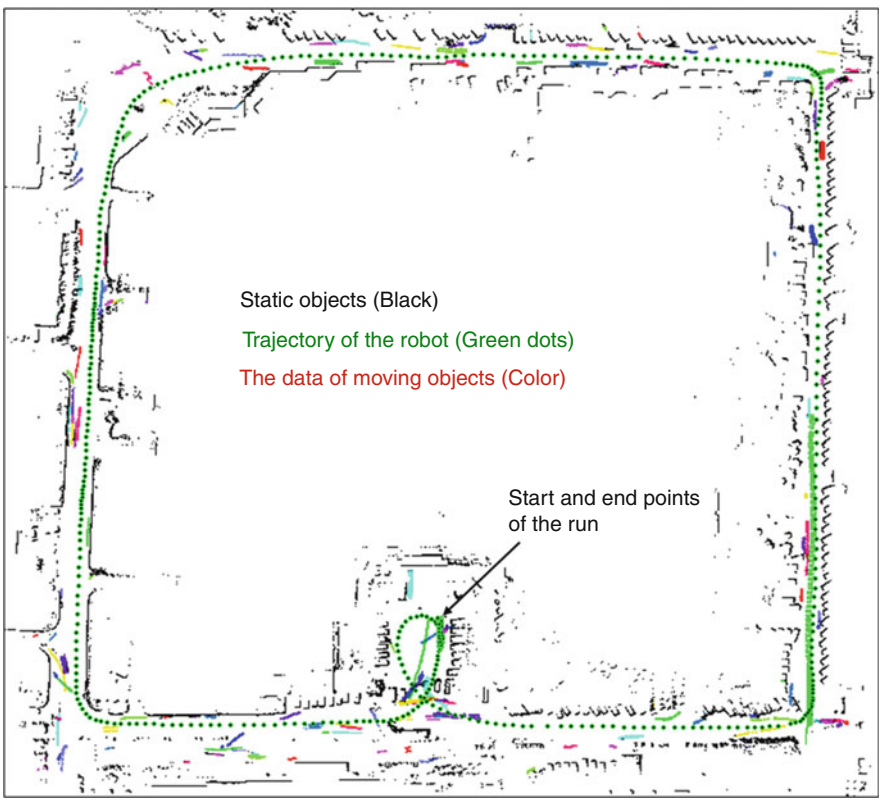
and all laser scan data, and then forwards the raw sensor logs to other modules through a wired local network, e.g., the data of L1 and GPS/IMU are forwarded to module 2, while those of L2–L5 are forwarded to processors 3 and 4.

In and urban environment, bridges, trees, tunnels, and buildings might obstruct GPS signals, and multi-path degrades the positioning accuracy of GPS/IMU based navigation units dramatically [43]. An algorithm of SLAMMODT was developed in the

authors' previous work [44], coupling a horizontal laser scanner with GPS/IMU, so that estimations of the robot's motion parameters can be refined by matching inter-frame sensing data to static objects (localization), meanwhile generating a consistent map in 2D grids, and detecting and tracking the data of dynamic entities at the horizontal level. The SLAMMODT is conducted in a horizontal level with an assumption that the ground surface is flat, and three kinds of outputs can be issued online-vehicle pose, a 2D grid map of static objects, and 2D trajectories of moving objects. A result after a cyclic run around a large building complex in a dynamic campus is shown in Fig. 5, where the course length is about 1 km.

In order to make collaborative usage of the data from different laser scanners, a calibration of sensor geometries is conducted in module 3 by referencing the sensor's coordinate system of L1. With the data of both sensor geometry and vehicle position, laser scans captured by different sensors at different vehicle positions can be integrated into a global coordinate system in module 4, so as to generate a map of the outside environment using the low-level primitive of 3D laser points. With such a representation, an operator can obtain visual knowledge of the environment, however, it is not informative for a robot, as such a map reflects only spatial existence through point sampling of object surface. In order for a robot to have semantic knowledge of the environment, such as objects, types, and their spatial relationships, a segmentation and classification of 3D laser points is conducted in module 5. A contextual map is finally reconstructed with each 3D object annotated, and the motion data of dynamic entities from processor 2 are integrated, which can be used to infer the relationships between different objects.

Currently, modules 1, 2, and 4 conduct online processing, module 3 is pre-processing directly after



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 5
A result of SLAMMODT after a cyclic run in a dynamic campus

each sensor system setting, and module 5 is an off-line procedure after a vehicle run. The algorithm details as well as experiments with modules 3, 4 and 5 are described below; module 2 is described in [44].

Sensor Calibration and Data Alignment

Literature Review

There have been dozens of intelligent vehicle systems using laser scanners for perception. For example, all 11 finalists in the DARPA Urban Challenge relied upon laser scanners (LIDARs) as their primary exteroceptive sensing modality [24, 25]. Laser scanners can provide angle and range measurements, which are the samplings on object surface, and can be easily converted into 2D or 3D coordinates. However, a single laser scanner might have limited viewing angles and occluded sensing space, e.g., the widely accepted SICK LMS 2** conducts 2D scanning on a plane within 180°. Therefore, it is important for a robot to possess a number of laser scanners to achieve a better understanding of the whole 3D environment. In order to integrate the data from different sensors, calibration is required to find sensor geometries, i.e., transformation parameters between different sensor's coordinate systems.

Normally, calibration methods are decomposed into intrinsic and extrinsic calibrations. Intrinsic calibration finds the local geometry of each sensor, such as the camera's focal length, optical distortion, principal point, and so on. These parameters can be estimated independently with other sensors. They experience only slight changes in temperature and humidity. Extrinsic calibration finds the relative geometry, e.g. rotation and transformation parameters, between different sensors that compose a rigid body. These parameters change during sensors' removal and set-up, and need to be calibrated after the sensors are mounted on the vehicle. Calibration of the sensor system using laser scanner(s) is poorly documented, except for a few papers on intrinsic calibration of laser scanners [45, 46], extrinsic calibration of laser scanners with camera [47–49], and multi-laser systems [50].

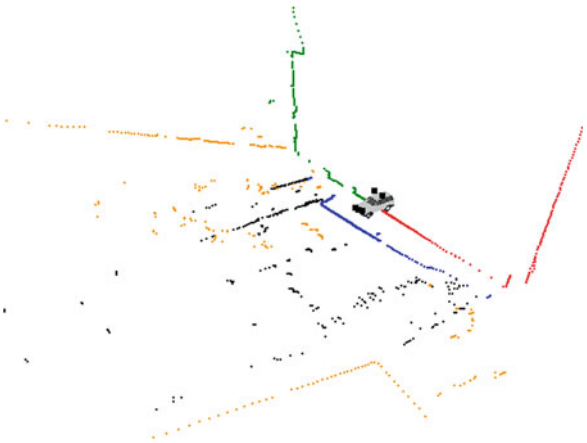
The procedure for extrinsic calibration, especially for a large sensor system, is notoriously labor intensive.

Relative geometry between different sensors' coordinate systems can be estimated by registering their data sets, and a relationship between the data spaces can be established. Registration of range data has been studied for decades [51–53], where sensor geometry is supposed to be fixed during the measurement of each set of range data. However, registration of mobile laser scan data is different from that of a stationary platform.

This research focuses on extrinsic calibration of a multi-laser sensor system that uses a number of single-row laser scanners (LIDAR). A calibration method has been developed that is not restricted to the sensor system shown in Fig. 1, but is adaptable to different sensor numbers and layouts. Here, it is necessary to stress that (1) the laser beam of the laser scanners (e.g., SICK LMS2**) is invisible (near infrared band); (2) a traditional checkerboard is not useful here, as a single-row laser scan shoot only one line on it; and (3) locating a reflective target to catch a laser beam is labor intensive, especially with laser beams of upward scanning. A calibration method is desirable if it (1) does not require special facility and (2) uses the objects in the scene as landmarks.

Problem Statement

Without loss of generality, a laser scanner's coordinate system is defined as follows (see Fig. 6). The origin is at the principle point of the laser scanner. Laser beam



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 6
Instantaneous laser measurements at vehicle body frame

scans start from the x-axis, moving toward the y-axis. The z-axis is vertical to the laser beam's scanning plane, and xyz-axes compose a right-hand coordinate system. A range point $p_i^r = (k_i,\ r_i,\ s_i)$ is measured by the laser scanner p at scan line number k_i (corresponding to and denoted as "time" below), with range distance r_i, and data number s_i. $(k_i,\ s_i)$ locates the pixel index in range image (see Fig. 2), the value of which is assigned by $\frac{r_i \times 256}{r_{max}}$. Let α_{res} be the angular resolution for range sampling, e.g., $\alpha_{res} = 0.5$ in this research, s_i is converted to a scanning angle by $\alpha_i = s_i \times \alpha_{res}$, p_i^r is converted into a 2D coordinate at laser scanner p's coordinate system, and is represented in the following form.

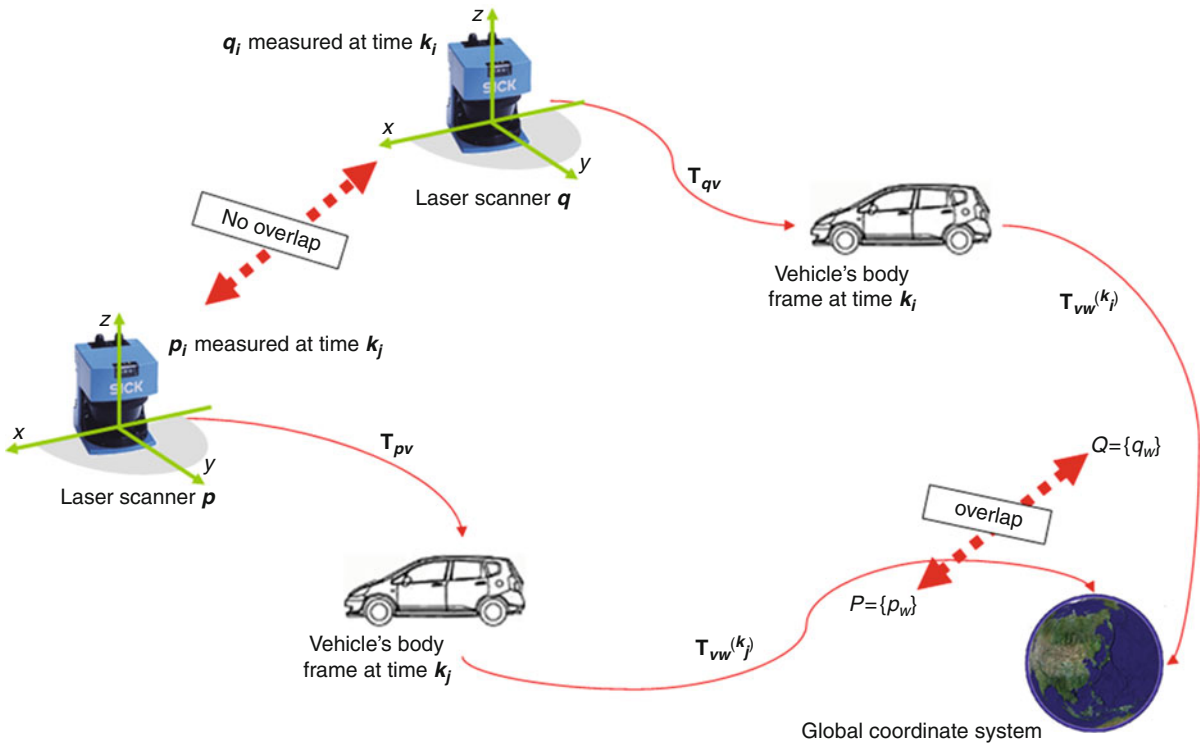
$$p_i = (r_i \cos \alpha_i,\ r_i \sin \alpha_i,\ 0, 1)^t$$

Let T_{pv} be the geometric transformation from laser scanner p to the vehicle body frame, which is defined on that of L1, p_i can be converted to the vehicle body frame as follows.

$$p_v = (x_v,\ y_v,\ z_v,\ 1)^t = T_{pv} \cdot p_i$$

Given an initial set of sensor geometry (T_{pv}s), the laser points that are measured by L1–L5 at a certain moment can be integrated into the vehicle body frame as shown in Fig. 7, where colors denote for different sensor data: light blue for L1, dark blue for L2, orange for L3, green for L4, and red for L5. It can be found that, at any moment, an instantaneous measurement obtains data only on five 2D planes, which do not compose a 3D measurement of the whole environment. Also, even T_{pv}s are erroneous, it cannot be discriminated from the instantaneous laser scan lines, as there are no or few intersections between the data spaces of different sensors.

On the other hand, with module 2 – SLAMMODT, the vehicle robot is able to build a 2D grid map of the environment, meanwhile it determines its location and orientation within the map as shown in Fig. 5. The map can be a relative one, if referencing to the vehicle body frame at the starting point, or a global one if incorporating GPS/IMU inputs, where a fiducial vehicle body frame is defined on that of L1. Position estimation finds the location and orientation of the vehicle body frame



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 7
Sensor geometry and data geo-referencing [42]

with respect to the reference one (denoted as global coordinate system), as a result, a transformation matrix T_{vw}^{ki} is obtained at each time instance ki, associating the vehicle's body frame to the global one. With both sensor geometry T_{pv} and vehicle position T_{vw}^{ki} at time k_i, p_i can be converted to the global coordinate system as follows (see Fig. 7). $P_w = \{p_w\}$ denotes the geo-referenced data set of laser scanner p.

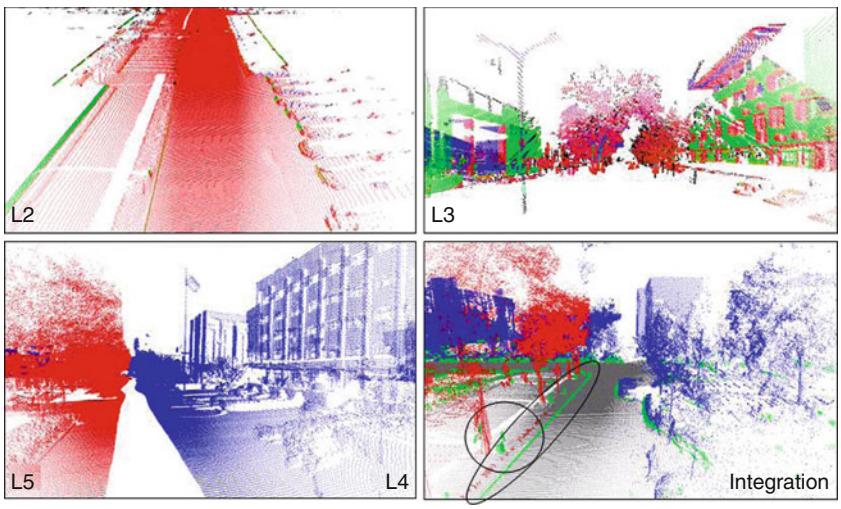
$$p_w = (x_w, y_w, z_w, 1)^t = T_{vw}^{ki} \cdot T_{pv} \cdot p_i$$

In the same way, let q be another laser scanner, q_i^r, q_i, q_w, and Q_w denote a range point, 2D coordinates at the sensor's coordinate system, a 3D point in the global coordinate system, and the geo-referenced data set of the laser scan measurements, respectively.

Figure 8 shows the geo-referenced 3D laser points, where the top-left, top-right, and bottom-left drawings are the data of each individual laser scanner, while the bottom-right is an integrated version. For the purposes of discrimination, colors in the drawings in the top row denote the local surface normal of the laser points, while colors in the drawings in the bottom row denote data from different sensors. It can be found that although the data generated by each individual sensor contains only partial knowledge of the whole environment, e.g., the data of L2 represents road surface only, the laser points have good consistency that describes object geometries in its vision field. However, when integrating the data from different sensors into a single one, displacement can be found in the duplicated measurements of common objects. It reflects that, although laser scanners have no intersection in their instantaneous measurement, a single object could be measured by different sensors at different time instances. When using the same vehicle position T_{vw}^{ki}s, the displacements come from the erroneous sensor geometries (T_{pv}s). This suggests that sensor geometries can be refined by minimizing the displacements between the overlapped measurements of different sensors after data geo-referencing, i.e., data registration. Let $T_{pq} = T_{pv}^{-1} \cdot T_{qv}$ denote the relative geometry between laser scanner p and q. Calibrating the sensor geometry between laser scanner p and q allows conversion into the following data registration problem:

$$T_{pq} = \arg\min_{T_{pq}} D(z(P_w), z(Q_w))$$

where $z(\cdot)$ is a function projecting P_w and Q_w to a matchable data space, and $D(\cdot\ ,\ \cdot)$ is a distance measure evaluating the displacement (i.e., distance) between two data sets. In solving the problem, a coarse-to-fine procedure is taken. Note that if q is L1, as the vehicle body frame is defined on L1, $T_{pq} = T_{pv}$.



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 8
Geo-referenced 3D points of each individual laser scanner and their integration [42]

Coarse Registration

A geometric transformation T_{pq} is first initialized according to hardware design, then refined through matching their data sets, i.e., P_w and Q_w. If a direct correspondence between a number of landmark points in P_w and Q_w could be established, T_{pq} can be solved using a Least Squares Method. However, when laser scanners face different directions, partial observations and occlusions make direct correspondence an extremely difficult issue.

Recall the problem statement in previous section. A fiducial vehicle body coordinate system is defined on that of L1. Sensor calibration can be stated more specifically to achieve an alignment of all other laser scanners (L2–L5) with reference to L1. With the output of module 2 – SLAMMODT, a 2D grid map generates a horizontal reference of the environment, while T_{vw}^{k} locates a vehicle at any time k on the map as shown in Fig. 5. In a calibration procedure, vehicle positions T_{vw}^{k} s are considered adequately accurate, leaving T_{pq}s for refinement. In this research, a two-step procedure is designed in coarse registration: horizontal and vertical matching, both of which generate projections to a 2D grid map on a horizontal level.

Horizontal Matching Horizontal matching generates a projection $z_h(\cdot)$, so that the laser points of vertical features, such as buildings, poles, and tree branches, are projected to a horizontal grid map. The grid value $v(\cdot)_{ij}$ can be 1 or 0, representing that the number of laser points projected to the grid cell is greater than or less than a predefined threshold. The reason for thresholding is to filter out irregular laser points and enhance vertical features. Horizontal matching is conducted by taking a laser scanner p as reference, adjusting T_{pq} to minimize the distance between $z_h(P_w)$ and $z_h(Q_w)$. A distance measure is designed based on the following two criteria.

$$C_1 = \sum_i \sum_j v(\cdot)_{ij} \rightarrow \min$$

C_1 means that the sum of all grid values should be minimized. As each grid cell has a value of either 1 or 0, C_1 means also that the number of grid cells that has a value equal to 1 should be minimized. Ideally, a vertical wall is converted to a line by $z_h(\cdot)$, a pole becomes a point and so forth. Laser points belong to the vertical features, such as walls or poles, tend to be projected to the limited number of grid cells corresponding to the lines and points. A concentrated projection is highly evaluated comparing to a distributed projection.

$$C_2 = \sum_i \sum_j v(p)_{ij} \cdot v(q)_{ij} \rightarrow \min$$

C_2 means that the sum of a production between the grid values $v(p)_{ij}$ and $v(q)_{ij}$ should be maximized. As each grid cell has a value of either 1 or 0, it also means that the number of matched grid cells that have a value equal to 1 should be maximized. A distance measure is designed by summarizing the above criteria.

$$D_h(z_h(p),\ z(q)) = C_2(p,q) - C_1(q) \rightarrow \min$$

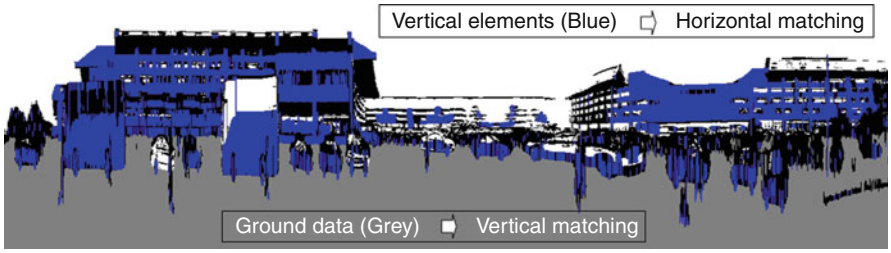
Vertical Matching Vertical matching generates a projection $z_v(\cdot)$, so that the laser points on ground surface are projected to a horizontal grid map. The grid value $u(\cdot)_{ij}$ represents the ground elevation at the grid cell, which is assigned by taking a mean of the z values of all laser points that are projected to the grid cell. A distance measure is designed as follows to evaluate the displacement between the elevation values:

$$D_v(z_v(p),\ z_v(q)) = \sum_i \sum_j |u(p)_{ij} - u(q)_{ij}| \rightarrow \min$$

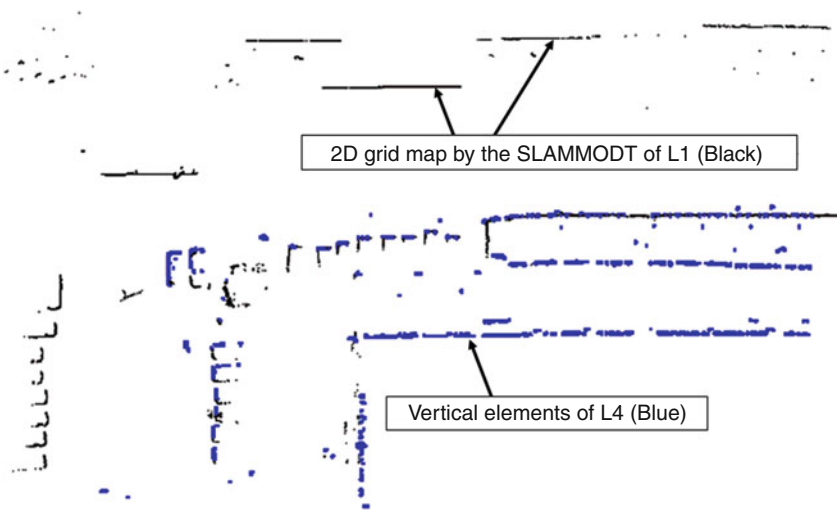
Figure 9 shows an example of extracting the horizontal and vertical elements from the data of L4. A local surface normal is estimated for each laser point using its neighborhood data on scanning order. Either the horizontal and vertical elements are extracted by thresholding on their local surface normal. The vertical elements are used for horizontal matching, where an example of registering L4 on the 2D grid map of L1 is shown in Fig. 10. The horizontal elements, i.e., ground data, are used for vertical matching, where an example of registering L3 with L4 and L5 is shown in Fig. 11. In this research, both horizontal and vertical matching is conducted by adjusting the parameters of T_{pq}, so as to minimize D_h and D_v, respectively, in an exhaustive way.

Fine Registration

In coarse registration, data are represented in 2D grid maps at a horizontal plane, where resolution in



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 9
Example of extracting horizontal and vertical elements



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 10
Example of horizontal matching of vertical elements from L4 and L1

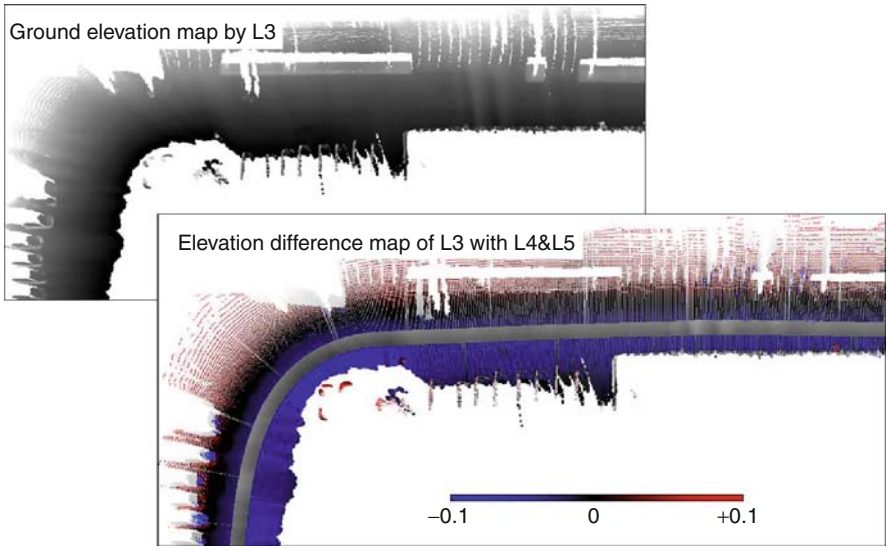
describing object details as well as registration accuracy are limited by such a representation. Find registration approach is to refine the coarse registration results.

Given the sensor geometries of T_{pv} and T_{qv}, for any p_w measured by laser scanner p, a mirror range point ${}^{q}p_j^r = (k_j, r_j', s_j)$ is estimated at the range image frame of laser scanner q. Let $Q^r = \{q_j^r\}$ denote the range image measured by laser scanner q, while $Q_p^r = \{{}^{q}p_j^r\}$ for the mirror range image generated from the data measured by laser scanner p. The objective of fine registration is to find a T_{pv} that minimizes the difference between Q^r and Q_p^r. It is defined as follows:

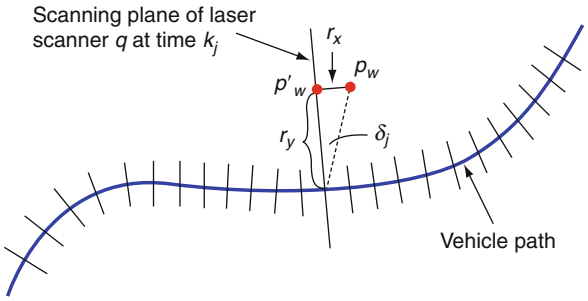
$$T_{pv} = \arg\min_{T_{pv}} D(Q^r,\ Q_p^r)$$

Given p_w, a mirror range point ${}^{q}p_j^r = (k_j, r_j', s_j)$ is estimated, where k_j is the number of a scan line at laser scanner q's data space, s_j is the data number at k_j, r_j' is the range distance that p_w be measured on the scanning plane of k_j at range beam angle $\alpha_j = s_j * \alpha_{res}$. A mirror range point is estimated in an order of $k_j \rightarrow s_j \rightarrow r_j'$, where estimation of each parameter, as well as a definition to $D(\cdot\ ,\ \cdot)$ is described below.

Estimation of k_j For any laser scan k_j at the data space of laser scanner q, an orthogonal projection is done from p_w to the scanning plane as shown in Fig. 12, where r_x is the orthogonal distance, r_y is the Euclidean distance from sensor q's viewpoint at k_j to the



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 11
Example of vertical matching of ground data from L3 with L4 and L5



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 12
Locating a scan line k_j at the data space of q for a laser point p_w

orthogonal point p_w', and $\delta_j = \arctan(\frac{r_y}{r_x})$. k_j is located as the laser scan of minimal δ_j and $\delta_j < \tau$, where τ is a threshold representing the diffusion of laser beam.

Estimation of s_j and r_j' With the orthogonal point p_w', a mirror point in sensor q's coordinate system is estimated as follows,

$$^{q}p_j = (^{q}x_i, {}^{q}y_i, {}^{q}z_i \approx 0, 1)^t = (T_{qv})^{-1} \cdot (T_{vw}^{kj})^{-1} \cdot p'_w$$

So that s_j and r_j' are estimated as

$$s_j = \arctan(\frac{^{q}y_i}{^{q}x_i}) \Big/ \alpha_{res}$$

$$r_j' = \sqrt{(^{q}x_i)^2 + (^{q}y_i)^2}$$

Definition of $D(\cdot\,,\,\cdot)$ Let

$$sign_{ij} = \begin{cases} 1 & if\ \delta_{ij}^{q} \cdot {}^{q}\delta_{ij}^{p} \cdot (||r_{ij}^{q} - {}^{q}r_{ij}^{p}|| < \varsigma \\ 0 & otherwise \end{cases}$$

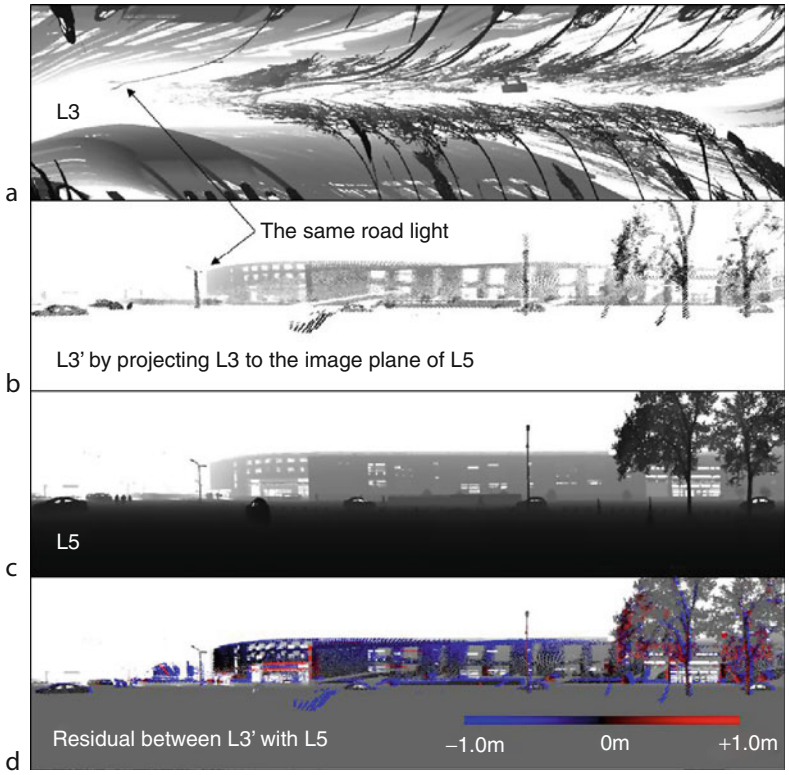
where δ_{ij}^{q} and $^{q}\delta_{ij}^{p}$ are binary values, denoting whether the range value r_{ij}^{q} at Q^r and $^{q}r_{ij}^{p}$ at Q_p^r, respectively, are valid (1 for valid). ς is a threshold, defining the bound for matched pair of range values. $sign_{ij} = 1$ denotes that the range points of Q^r and Q_p^r at pixel index (i,j) are matched ones. $D(\cdot\,,\,\cdot)$ is designed based on the following two criteria:

$$C_3 = \sum_{i=0}^{W}\sum_{j=0}^{H} sign_{ij} \rightarrow \max$$

$$C_4 = \sum_{i=0}^{W}\sum_{j=0}^{H} sign_{ij} * ||r_{ij}^{q} - {}^{q}r_{ij}^{p}|| \rightarrow \min$$

$$D(Q^r, Q_p^r) = C_4/C_3 - C_3$$

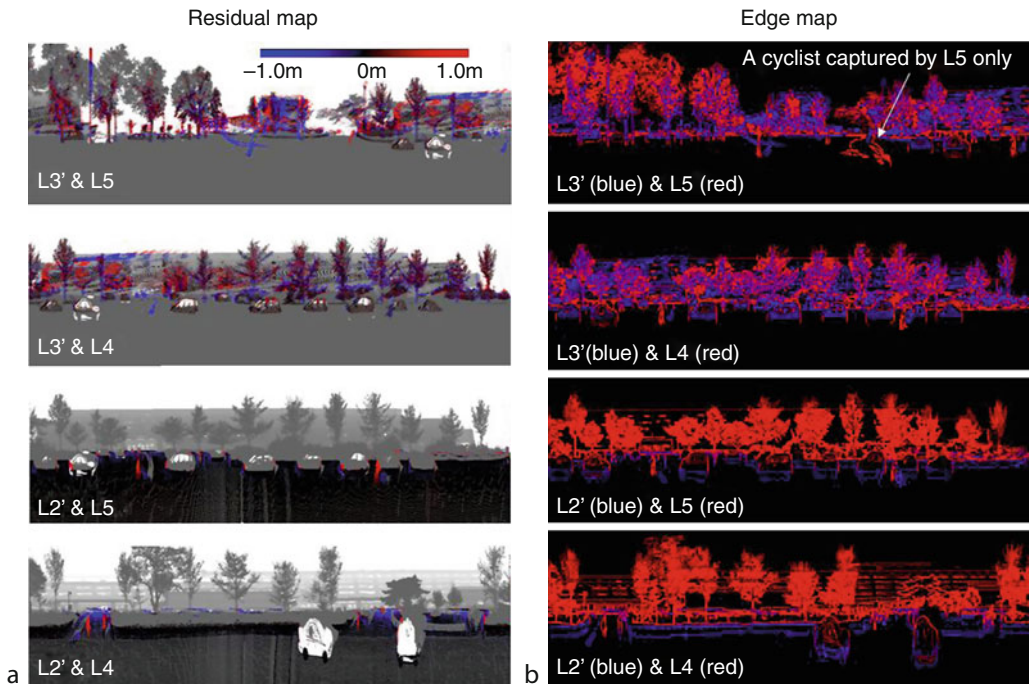
where C_3 counts for the number of matched range points, and C_4 is the sum of residuals between the matched range values. $D(\cdot\,,\,\cdot)$ is designed to maximize the number of matched range points, while minimizing the average of residuals between matched range values.



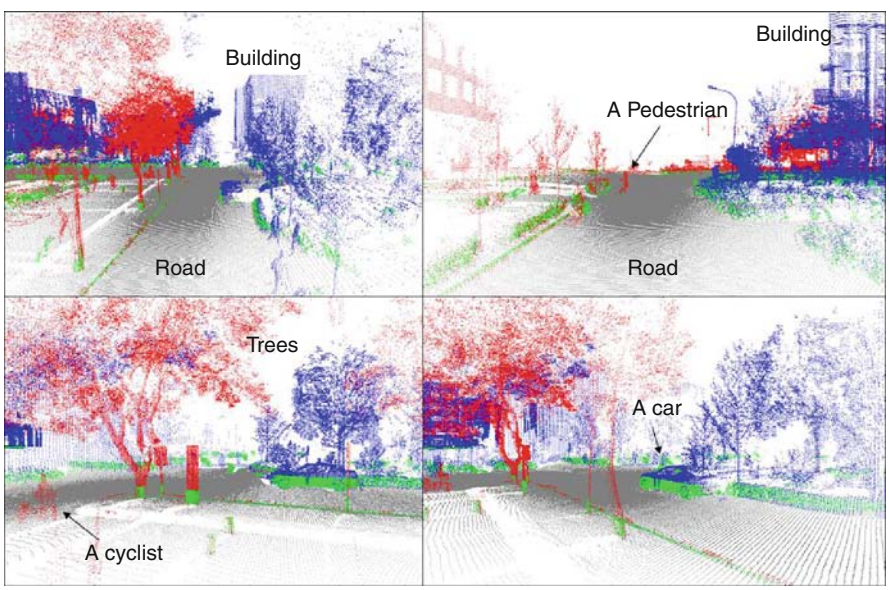
Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 13
Explanation of the fine registration procedure using experimental results

Figure 13 explains the procedure of fine registration using experimental results, where the data of L3 (a) is projected to the range image frame of L5, and a new range image L3′ is generated as shown in (b). (c) shows the original range measurement by L5. A match is made between L3′ and L5, as shown in (d), where residuals between corresponding range points are represented in a color band of [blue-black-red], while the range points of L5 that do not have matched pair in L3′ are shown in grey. Some other results are also shown in Fig. 14, which are represented in both the residual maps between mirror and original range images, and their edge maps. In the edge map of L3′ and L5, it can be clearly found that a cyclist is captured by L5, however not in L3′. This might happen in the measurements to moving objects, where due to the difference in sensors' viewpoints, a single object (place) might be measured by sensors at a different time period. Although the cyclist is captured by L5, when L3 measured the place, it was not there, and L3 failed to capture the cyclist all the time due to its high dynamic motion. On the other hand, this phenomenon suggests that by matching the data from different sensors, the dynamic factors in the environment can be detected. Due to occlusion, viewpoint difference, and the dynamic entities of the environment, matching between the mirror and original range image is conducted manually in this research, where an automated method is to be developed through future research.

After sensor alignment, the laser points from different laser scanners can be integrated into a global coordinate system with more consistency, as shown in Fig. 15, where the ground data of L2 is shown in gray, non-ground data of L2 containing those of road edges are shown in green; L4 is in blue and L5 in red. It can be found that good consistency between the data of different sensors is achieved through a calibration procedure, and an omni-directional measure of the whole environment is achieved through the collaborative exploration of multiple single-row laser range scanners.



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 14
Some results of fine registration. *Left column*: residual map between mirror and original range images; *Right column*: edge points extracted from mirror (*blue*) and original (*red*) range images



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 15
3D map of the environment created by integrating multi-laser sensing data [42]

Contextual Map Generation and Representation

An environmental map can be generated directly using the low-level geometric primitive of 3D laser points. However, such a map has limited capacity in representation, as it describes only spatial existence. An operator can easily understand from the data where objects are, and what kinds of objects they are, while a robot cannot. In order for a robot to have semantic knowledge of the environment, such as objects, types, and their spatial relationships, an automatic technique of converting a low-level map representation into a high-level one is important.

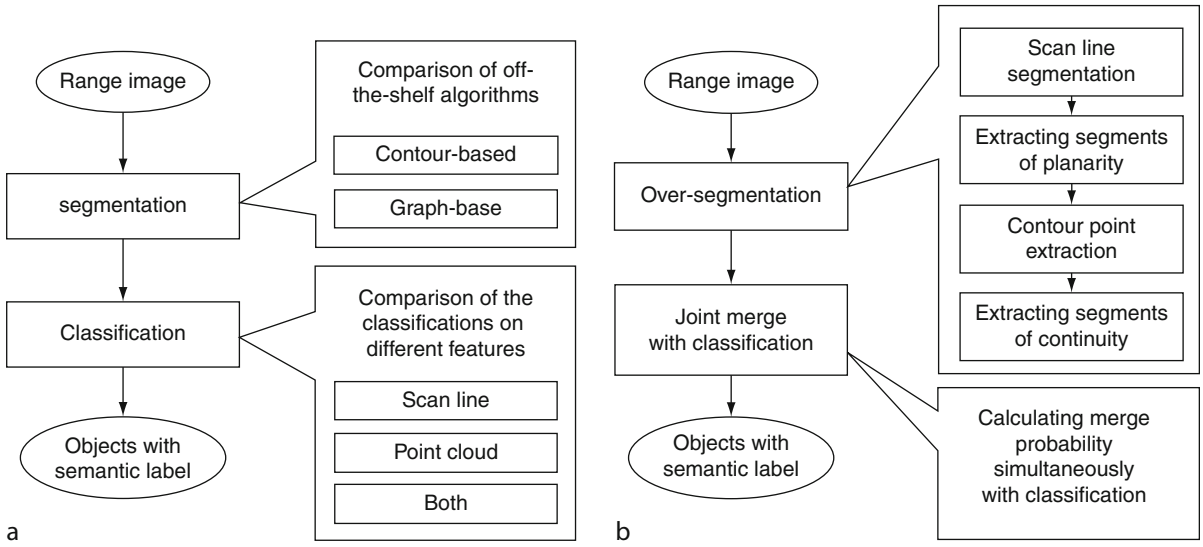
Literature Review

Through a bottom-up procedure, laser points can first be processed to find data clusters, i.e., laser points that are most likely to be the measurements of the same objects. They can then be recognized as certain kinds of objects, e.g., planar surfaces [54], line feature objects [55], cars [56], or natural objects [57]. In order to tackle the large number of laser points, [58, 59] tessellated the 3D space and projected laser points to voxels. The sequence of laser points can also be used as a whole to represent the geometric appearance of local surroundings, and a higher level knowledge of the robot's concurrent location, such as a doorway, corridor, or room, can then be inferred through machine learning techniques [60]. In addition, each individual laser measurement is considered to be dependent on its neighborhood in [61]. Their relationships are modeled using Markov networks, where labeling of each laser point is influenced by the labeling of others in its local vicinity. [62] also associated image cues with each laser point and varied the probabilistic framework using Conditional Random Field. Reference [63] further extended the method of [61], so that each node in the Markov network corresponds to a data patch, i.e., a superpixel (image patch) with corresponding laser points in successive scans, rather than a laser point. In addition, a recent report can be found in [63], where an aerial laser scan data is processed to label the small objects, such as posts, lights, and cars, in an urban environment. The problem is solved by localization, segmentation, representation, and classification procedures in a sequential way.

On the other hand, laser scan data can be represented in the form of a range image, where each pixel represents a depth value, and its index corresponds to the sequential order of measurements. Thus, beam origin and angle of each depth value can be retrieved, and depth value can then be easily extended to a 3D coordinate. There is a large body of work addressing range image segmentation. A famous report comparing the major segmentation methods can be found in [64]. Many of the methods are motivated by the needs for recognizing industry parts [65] or registering the data taken at different locations [66]. These works always assume simple or well-defined object geometry. There are still a few research works processing range images of real-world scenes. [67] considers a real-world indoor and outdoor scene by modeling the man-made objects using planes and conics, modeling free-form objects using splines, and modeling trees using 3D histogram; segmentation and model fitting for each segment is formulated in a data-driven Markov Chain Monte Carlo procedure.

Image segmentation and semantic interpretation have been studied extensively in the field of computer vision. As the data form of a range image is consistent with that of a visual image, many methods developed in the field of computer vision are of great reference for the processing of range image [68–71].

As shown in Figs. 2, 8, and 15, laser scan data can be represented in the forms of 2D range images or 3D laser points, where for a 2D range image, each pixel can be converted into a 3D point in a global coordinate system as addressed previously. In this research, range image is chosen as an interface for data representation. More specifically, this research concerns the segmentation and classification of range images in the real world, however estimations are conducted by retrieving the 3D coordinates of each range pixel. In dealing with the problem, a sequential flow as shown in Fig. 16a is generally used [72], where classification is conducted subsequently to segmentation. The framework is straightforward and easy to implement. However, when facing a complex environment with different kinds of objects, segmentation failures might happen due to the heterogeneous geometric properties between different kinds of objects. Unified frameworks are also proposed [68–71] that cope with the segmentation and classification problems simultaneously (see Fig. 16b).



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 16
Sequential (**a**) and simultaneous (**b**) processing flows for range image segmentation and classification

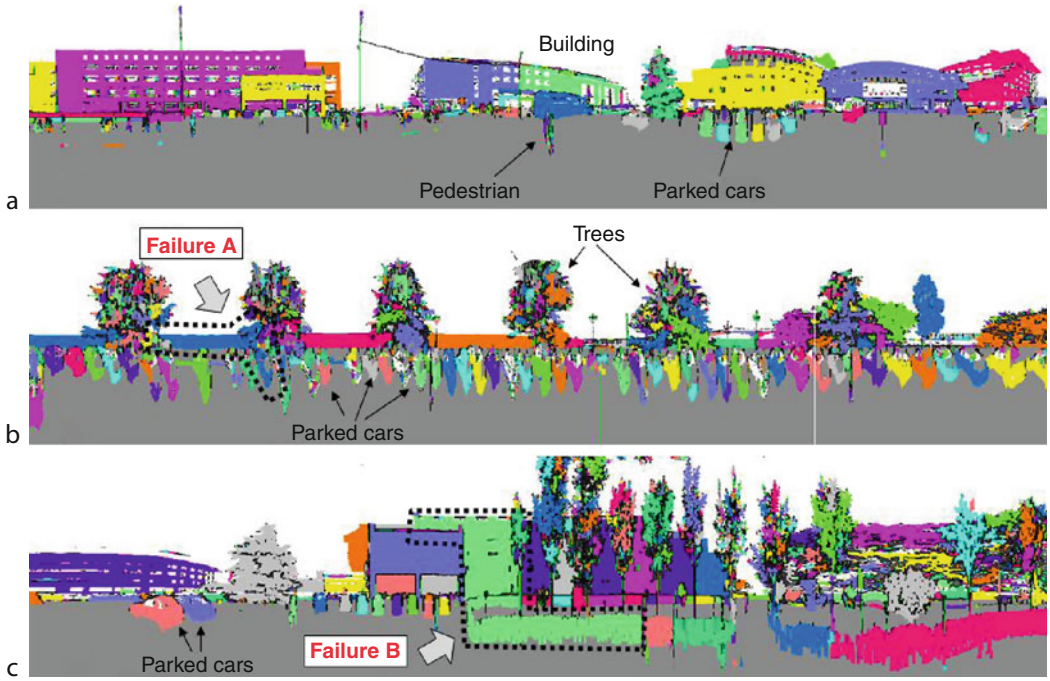
This research examines the performances of both sequential and simultaneous processing flows in solving the problem of segmentation and classification of the range images in complex outdoor environments. The algorithmic and implementation details as well as experimental results are described below.

Segmentation

First, the data of the ground surface are discriminated and removed from the range image, so making the data of vertical objects spatially disconnected. Ground elevation at the point of host vehicle can be assumed, as the vehicle stands directly on the ground, and the sensor-vehicle is composed of a rigid body. For each laser scan line, segmentation is conducted, and the nearly horizontal line segment that has at least one end point within the bound of assumed ground elevation is extracted as the ground data, as the grey points in Fig. 17.

Contour-Based Segmentation In segmenting the rest of range image, contour points are first extracted, where the Euclidean distance between the 3D coordinates of each contour point with its four neighbors in range image is larger than a given threshold; a point-based region growing is then conducted to extract the data segments with spatial continuity (different color in Fig. 17 denotes a different segment). The method is efficient in extracting the data segments of most of the objects, and they are spatially disconnected. As shown in Fig. 17, although over-segmentation (i.e., a single object can be separated into a number of segments) occurs, especially for trees and pedestrians, most of the segments contain the data of a single object; with under-segmentation, different objects are merged into a single segment. Failure A merged the data of a piece of wall and a nearby car. Failure B merged the data of building surface and a clump of bush, where in some places, they are spatially close.

Graph-Based Segmentation Data acquired by laser scanner usually spatially connect with its neighbors in a range image. This property suggests a graph-based segmentation. The set of pixels can be represented as a weighted undirected graph $G = (V, E)$, where V is the set of nodes standing for pixels and E is the set of edges between pixels and their neighbors in 3D space. Thus, segmentation can be formulated as the problem of cutting an undirected graph G, where many mature algorithms exist [73, 74]. This research chose the FH algorithm [74] for its good performance in catching the

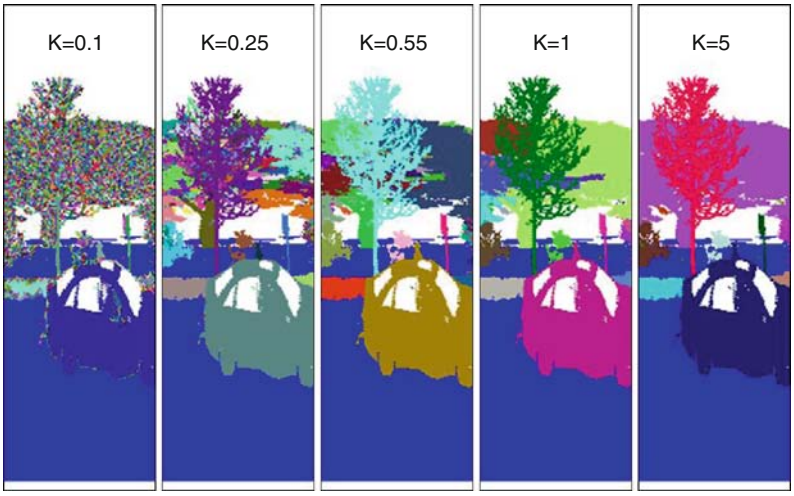


Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 17
Results of contour-based segmentation

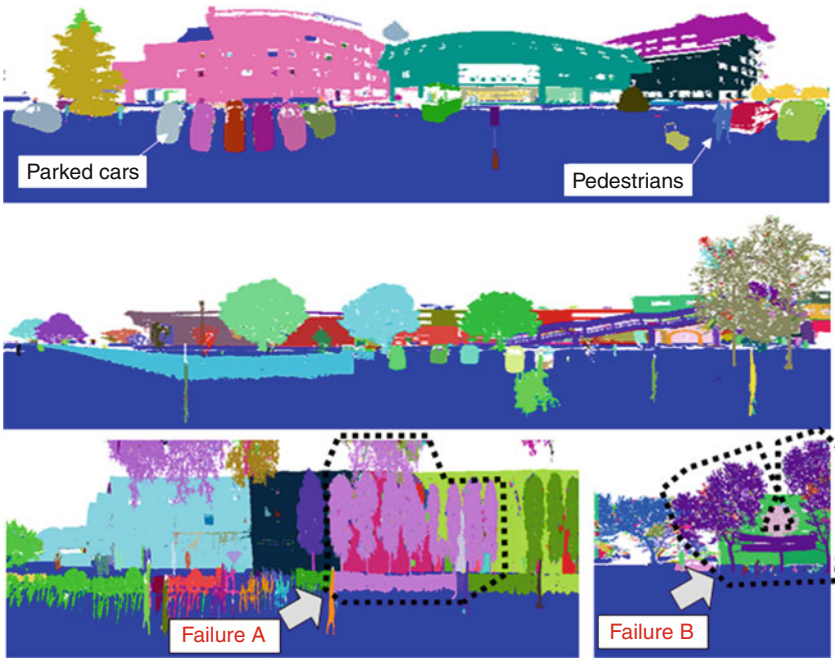
non-local properties and adaptability to online procedure. A range segment s is composed of a set of range points as well as weighed edges that represent the spatial relationships between neighbor range point pairs. The weight $w$of an edge $e$is assigned by the Euclidean distance between the 3D coordinates of two neighbor range points, which represents their spatial continuity. For any segment s that has ns range points, an internal difference $(Int(s) = \max_{e \in s} w(e))$ as well as a regulation term $(\tau(s) = \frac{k}{ns})$ is defined, where k is a predefined experimental factor. For any pair of segments (s_a and s_b), a minimum internal difference $MInt(s_a, s_b) = \min(Int(s_a) + \tau(s_a), Int(s_b) + \tau(s_b))$ is defined. The pair of segments s_a and s_b can be merged if and only if it has $MInt(s_a, s_b) < w(e_{a,b})$, where $e_{a,b}$ is an edge that has its two end points in s_a and s_b, respectively. Experimental results reflect that the parameter setting of k in the regulation term $(\tau(s) = \frac{k}{ns})$ is crucial to the segmentation result as described in Fig. 18. A small k causes objects to be over-segmented, while enlarging $k$might cause the data of different objects to merge. Figure 19 shows some of the results with $k = 1$. It can be found that the problem of over-segmentation is reduced compared with the results of a contour-based one, while under-segmentation is even heavier. Failure A merged the data of a number of trees and a clump of bush. Failure B merged the data of a banner and two neighboring trees.

Classification

A segment contains a set of 3D laser points, which describes the geometry of an object surface through the representation of point clouds. It is straightforward to extract data cues from the set of 3D laser points, and devise a classifier by evaluating their classification properties. On the other hand, a single-row laser scanner measures the environment in a mode of scan-line by scan-line. A scan-line measurement on a planar object, such as the surface of a building or road, can be modeled using a line segment. The measurement of a free-form object, such as a tree, yields many small line segments and isolated points. An isolated point can be considered as a specific type of line segment. A segment



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 18
Parameter setting in graph-based segmentation



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 19
Results of contour-based segmentation

can also be considered as a set of scan line segments [75, 76], where data cues are extracted and evaluated to compose a classifier. Of course, a classifier can also be devised by a combinatory evaluation on the properties of 3D laser points and line segments. In this research, three classifiers are developed on line segments, 3D laser points, and their combinations, which are described below as well as an experimental comparison.

Classification on Scan Line Segments Let L denotes the set of object classes, i.e., $L = \{l|building, road, tree, person, car \cdots\}$, y be the label of a segment s,

and yk be the label of the kth line segment. Given data cue I, classification of segment s on scan line segments is formulated as follows.

$$y = \bar{l} = \arg\max_{l} \prod_{k} P(yk = l|I)$$

Let $I = \{d_1, d_2 ...\}$, the probability of a line segmentlabeled l, i.e., $P(yk = l|d_1, d_2, ...)$ can be further extended according to Bays' rule (i.e., a naïve Baysian classifier).

$$P(yk = l|d_1, d_2, ...) = P(yk = l) \prod_{i} P(d_i|yk = l)$$

where $P(yk = l)$ is a prior knowledge of a scan line segment labeled l. It is initialized as an equal distribution and updated after each iteration according to the percentages of data labeling.

$P(d_i|yk = l)$ is the likelihood measure (denoted by $\lambda_l^{yk}(d_i)$), when given a label $yk = l$, the probability d_i of the scan line segment can be observed. The data cues extracted from each scan line segment are listed in Table 1. For any pair of data cues d_i and object label $yk = l$, a likelihood measure $\lambda_l^{yk}(d_i)$ is trained using a set of manually labeled scan line segments. A histogram is first generated on data samples, then Gaussian fittings are conducted on each distinctive picks, followed by a normalization so that integration of the graph is 1. The likelihood measures trained in this research are shown in Fig. 20.

Classification on 3D Laser Points Let Y denote the label on the cloud of 3D laser points. Classification is formulated as follows:

$$y = \bar{l} = \arg\max_{l} P(Y = l|I)$$

Data cues extracted from a cloud of 3D laser points are listed in Table 2. Differing from that of scan line segments, classification of 3D laser points does not evaluate the properties of each individual laser point but treats them as a whole. The likelihood measures $\lambda_l^{yk}(d_i)$ of a certain pair of data cues and object labels are generated as demonstrated in Fig. 21. However, due to the limited and unbalanced number of sample data (see Table 3), a naïve Baysian classifier did not yield satisfying results. In classification of 3D laser points, an off-the-shelf method, LIBSVM [77], is used.

Classification of Both 3D Laser Points and Scan Line Segments A classifier of both 3D laser points and scan line segments is designed through the multiplication of probabilistic estimations on each individual primitive, as follows:

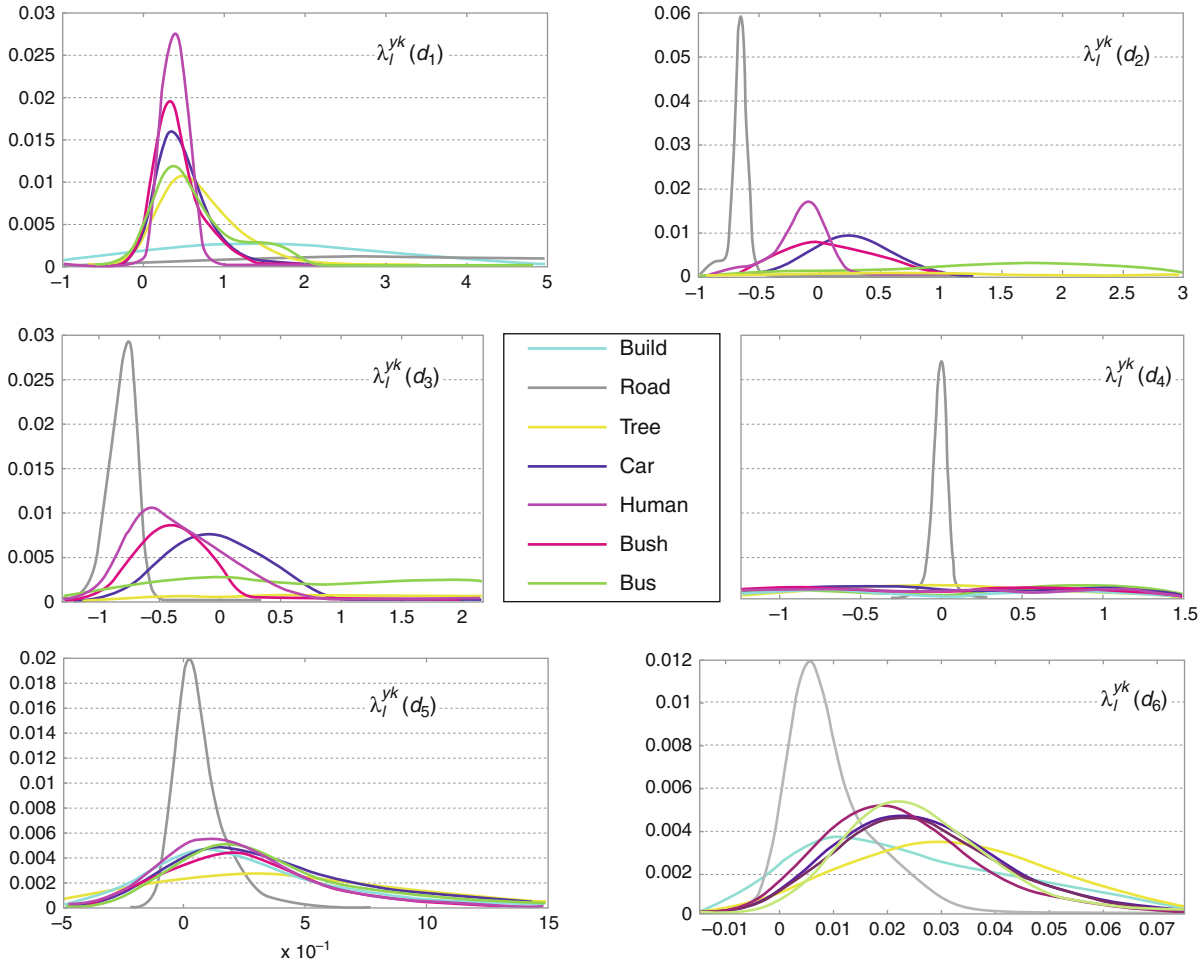
$$y = \bar{l} = \arg\max_{l} P(Y = l|I) \cdot \prod_{k} P(yk = l|I)$$

Comparative Experimental Results In training the classifiers, an interactive tool is developed to generate training data samples. A range image is first processed to find local surface normal for each laser point, edge point that has discontinuous change in range value, scan line segments, super segments, etc. These results are compared by an operator to discriminate the boundaries of individual objects. For each object, the operator will manually draw a boundary and assign a label. The software will output the scan line segments that are inside the boundary, as well as the label, for training classifier $P(yk|I)$. The software will also output all laser points within the boundary, as well as the label, for training classifier $P(Y|I)$.

In experiments, two simultaneously measured range images from sensors L4 and L5 are manually labeled. The manually labeled data of L4 is used in training classifiers. The number of training samples with respect to each object class and classification method are listed in Table 3. Three classifiers are trained on the features of scan lines (classifier 1), 3D laser points (classifier 2), and their combination (classifier 3). On the other hand, the range image of L5 is first segmented then classified, and the result is compared with those of manually labeled

Dynamic Environment Sensing Using an Intelligent Vehicle. Table 1 Data cues of a scan line segment

Feature	Definition
d_1	Length of the scan line segment
d_2	Maximal height value
d_3	Minimal height value
d_4	Z-coordinate of the directional vector
d_5	Mean of line regression
d_6	Variance of line regression



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 20
Likelihood measures for the classification of a scan line segment [41]

ones. Here, contour-based segmentation is exploited, and the segments are classified using each of the classifiers for comparison. A set of results is shown in Fig. 22, where a number of objects are highlighted. Object 1 includes cars that are correctly classified by all classifiers. Object 2 is a bus that is correctly classified by classifier 1 and 3, but failed in classifier 2. A reason might be the limited number of training samples. Objects 3 and 4 are pedestrians, which are failed in all three classifiers, pointing out that classification performance related to pedestrians needs to be improved through future work. Object 5 is a small house, the shape of which is similar to a bus. Classifier 1 misclassified it into a bus, while two others answered correctly.

Simultaneous Segmentation and Classification

As discussed previously, no matter whether it is a contour-based segmentation or a graph-based one, the same problem occurred. The balance between over-segmentation and under-segmentation can be guided through a parameter setting that is universal for all kinds of objects.

Segmentation is making a partition in data, where in each partition cell (i.e., segment), data has the property of certain homogeneity. However, in facing a complex environment that contains different kinds of objects, different kinds of objects might have different homogeneities. A segmentation method that applys

Dynamic Environment Sensing Using an Intelligent Vehicle. Table 2 Data cues extracted from a cloud of laser points

Feature	Definition
d_1	Minimal height value
d_2	Maximal height value
d_3	Ratio of boundary point number versus total point number
d_4	Mean of height values
d_5	Variance of a histogram distribution on normal vectors
d_6	Major picks of a histogram distribution on normal vectors
d_7	Ratio of width versus length

varied criteria for different object classes is required, where the segmentation and classification issues are considered in a unified form.

The processing flow is shown in Fig. 17b, where a super-segmentation is conducted using the bottom-up heuristics, such as scan-line segments and contour points at different scales, and then the super-segments are merged considering the factor of object class, which is modeled as a problem of joint merger with classification. In super-segmentation, scan line segments are first extracted [69, 70] and merged according to their planarity; a region growing is then conducted to merge the left points and isolated scan line segments according to their spatial connectivity, stopping when a contour point is met. Super-segmentation requires that each segment be a partial observation of an object; it should not be a mixture of different objects. Therefore, strict criteria are used in the above region growing procedures. In a merge procedure, a key problem is defining a model that evaluates the homogeneities for a variety of data segments. The probabilistic formulation is described below.

Probabilistic Model Let s_i and s_j be a pair of neighboring segments with the label of y_i and y_j, respectively. Let s_{i+j} denote the merged segment of s_i and s_j, and y_{i+j} for its label. The probability for s_i merged with s_j, i.e., $P(s_{i+j}|I)$, can be estimated as follows;

$$P(s_{i+j}|I) = \sum_{l \in L} P(y_i = l|I) P(y_j = l|I) P(s_{i+j}|y_{i+j} = l, I)$$

Where, given data cues I, $P(y_i = l|I)$ and $P(y_j = l|I)$ evaluate the probabilities of s_i and s_j labeled to l, respectively. Here, the classifier of both 3D laser points and scan line segments is exploited. $P(s_{i+j}|y_{i+j} = l, I)$ is a likelihood measure, where, given the knowledge of object class $y_{i+j} = l$ and data cues I, the probability of s_i and s_j are the measurements to a single object, i.e., s_i and s_j are merged to s_{i+j}. The likelihood measures with respect to different kinds of objects are defined below.

Definition to Likelihood Measures As the object class is given, an evaluation based on a prior knowledge of the object class is required. They are defined experimentally below.

1. y = building or road :

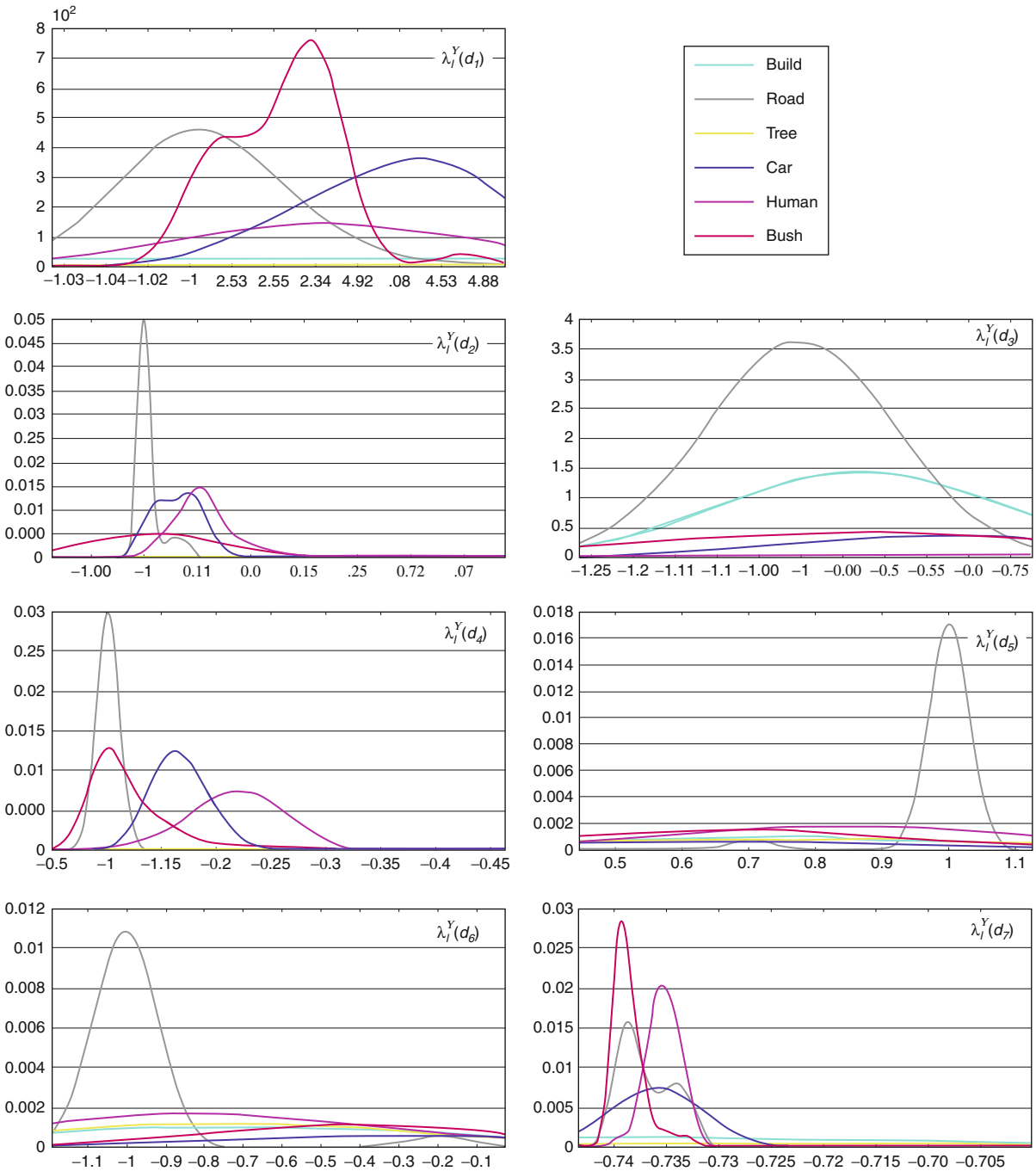
   Normally only a partial surface of building or road is measured, which could be modeled using a planar surface, with a certain volume (e.g., $\varepsilon = \pm 20$ cm) representing the errors in laser range measurements and modeling generalization. Let S denote the total number of laser points in the range segment, N the number of laser points within the volume, and α the angle between the surface normal and a vertical normal vector $(0, 0, 1)^t$. The likelihood measure is define

$$P(s|y = building,\ I) \propto \frac{N \times \sin\alpha}{S}$$

$$P(s|y = road,\ I) \propto \frac{N \times \cos\alpha}{S}$$

2. y = car:

   Normally data of a car can be restricted within a cube, so that τ_W, τ_L, and τ_H are defined according



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 21
The likelihood measures for classifying a cloud of laser points [41]

to the largest width, length, and height of a normal car. A cubic model is used to fit on the laser points of the segment and obtain a width (W_f), a length (L_f), and a height (H_f). A step function is defined (see Fig. 23)

$$y = f(x,\ \alpha_1,\ \alpha_2)$$

and

$$\int_{\max(\alpha_1-\varepsilon,0)}^{\alpha_2+\varepsilon} y\ dx$$

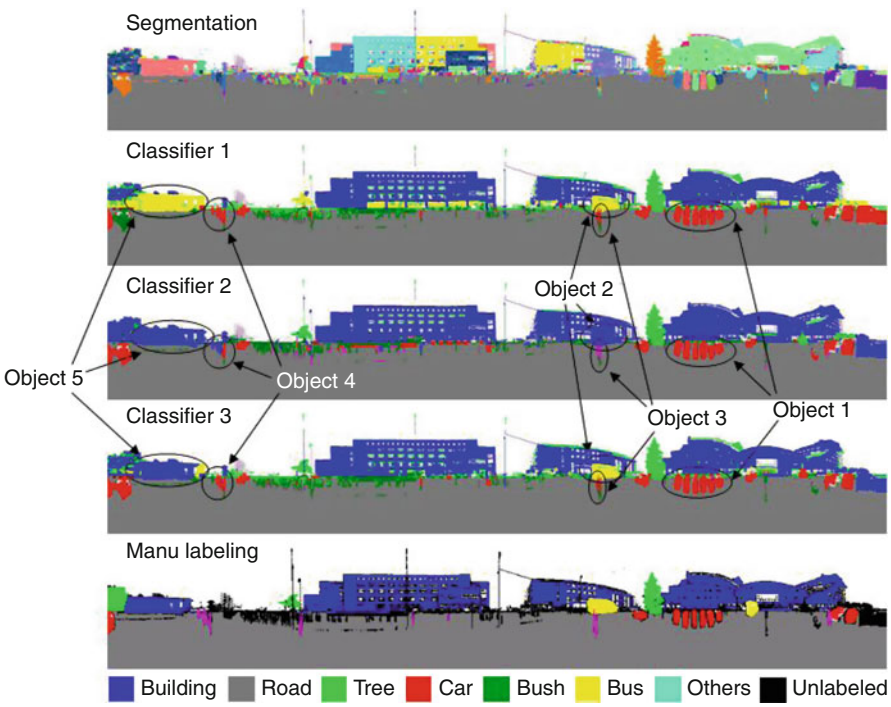
The likelihood measure is defined

$$P(s|y = car,\ I) \propto f(W_f, 0, \tau_W) \times f(L_f, 0, \tau_L) \times f(H_f, 0, \tau_H)$$

3. y = bus:
   Similar with that of a car.
4. y = person:
   Normally data of a person can be restricted within a cylinder, so that a radius (τ_r) and a height threshold (τ_h) is defined according to the size of a normal person. A cylindrical model is used

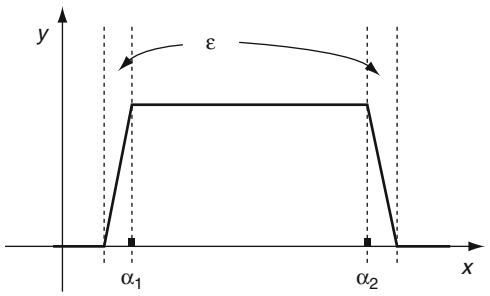
Dynamic Environment Sensing Using an Intelligent Vehicle. Table 3 The number of sample data in training for classifiers

Class	Scan line segments	Segments of 3D Laser points
Building	9,394	96
Road	10,714	23
Tree	4,122	148
Car	6,080	41
People	394	120
Bush	1,176	39
Bus	253	1
Total	32,133	468



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 22
Comparative results of different classifiers

to fit on the laser points of the segment and obtain a radius (r_f) and a height (h_f). The likelihood measure is defined



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 23
A step function for likelihood evaluation [41]

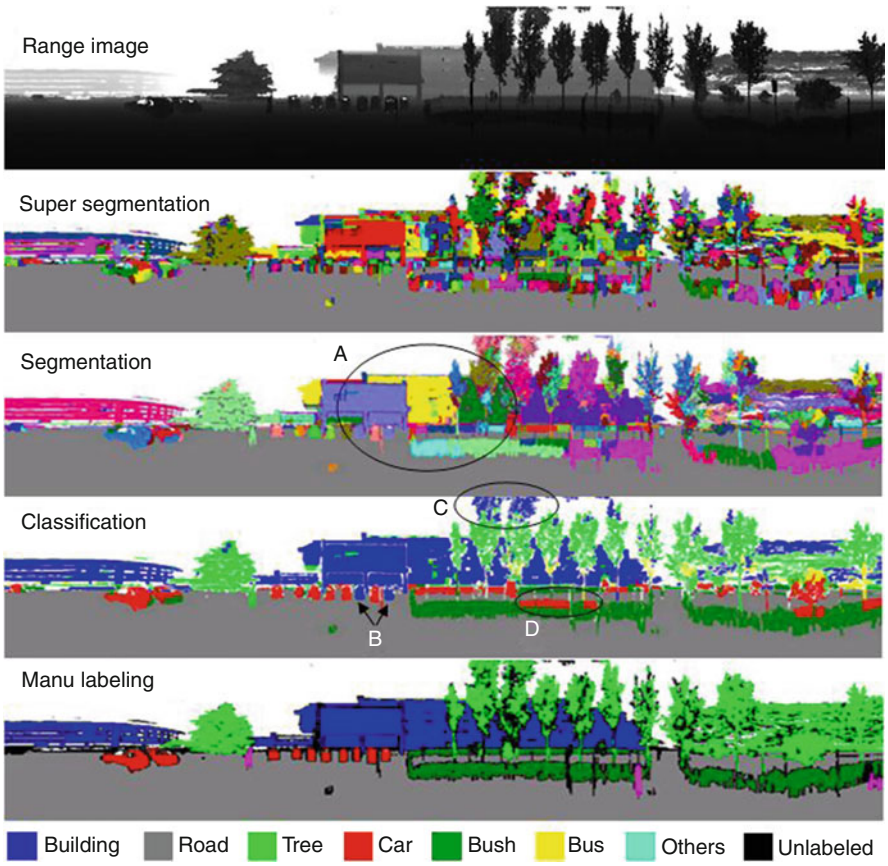
$$P(s|y = person,\ I) \propto \frac{\tau_r}{\max(\tau_r, r_f)} \times \frac{\tau_h}{\max(\tau_h, h_f)}$$

5. y = tree:
   Normally the range segment of a tree consists of many small line segments and edge points. Let S denote the total number of laser points of the range segment, and E the number of laser points on scan line segments. The likelihood measure is defined

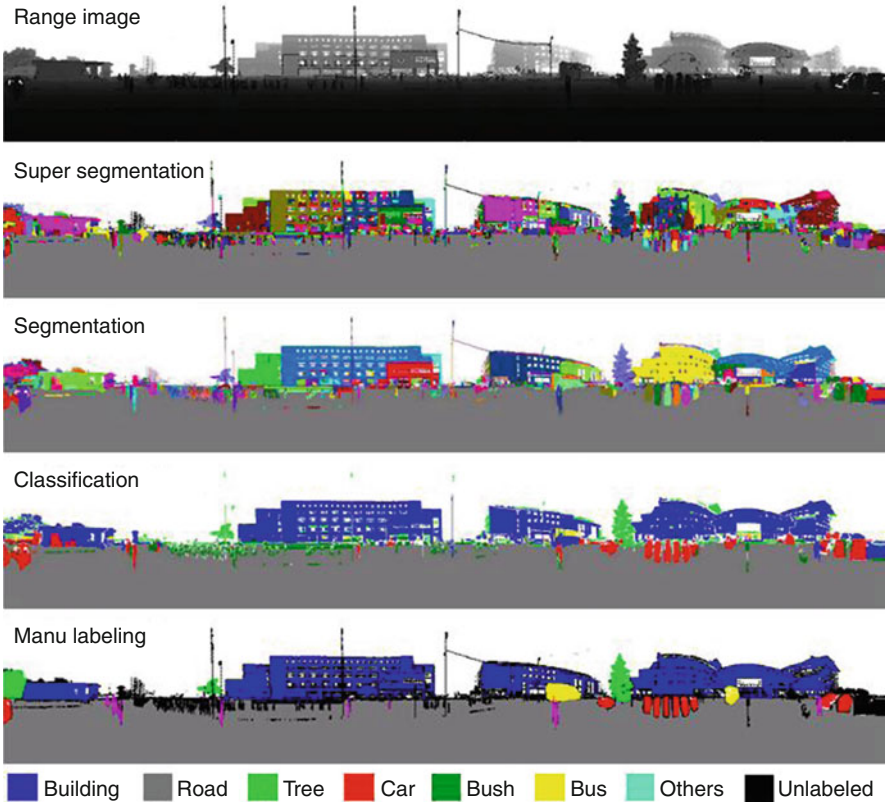
$$P(s|y = tree,\ I) \propto \frac{E}{S}$$

6. y = bush:
   Similar with that of a car or bus.

Experimental Results Figures 24 and 25 show experimental results of simultaneous segmentation and classification, where in each figure, the first row is an



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 24
Results of simultaneous segmentation and classification

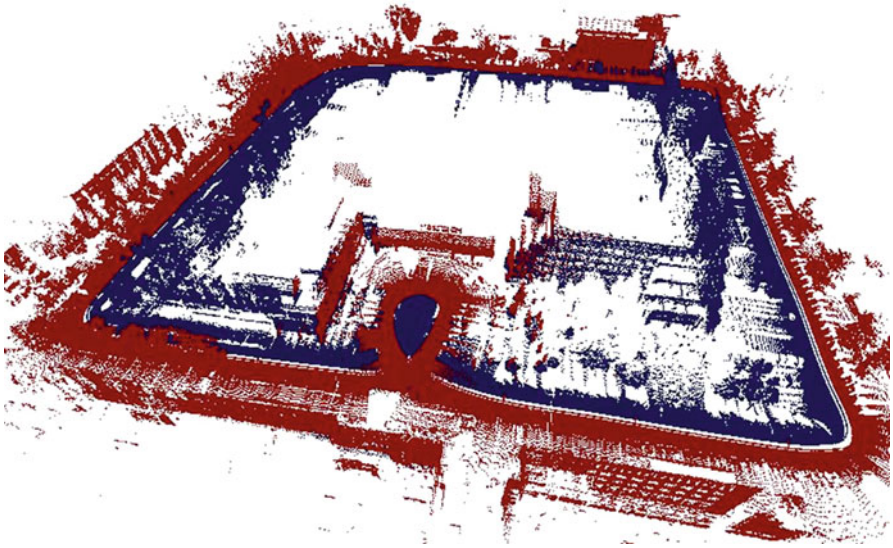


Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 25
Results of simultaneous segmentation and classification [41]

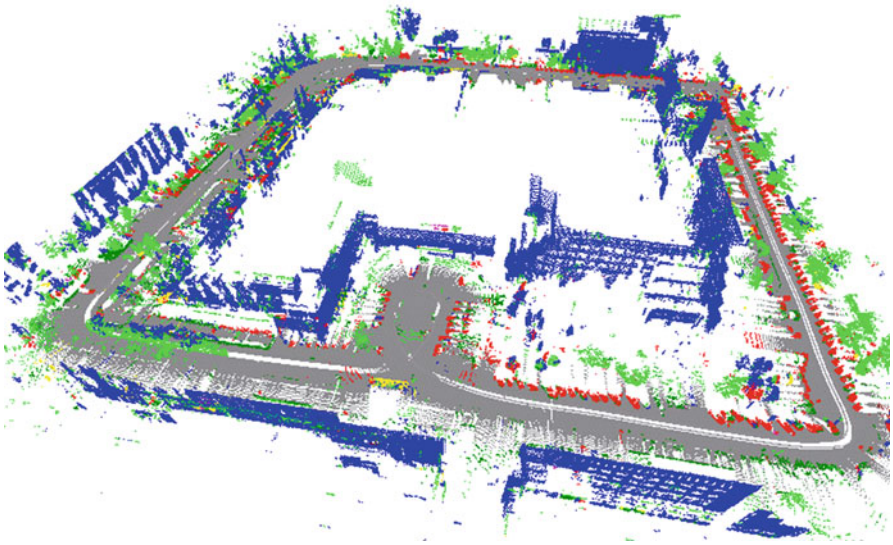
original range image, and the second row shows the segments after super-segmentation. The third and fourth rows are simultaneously obtained results, which are the segments after merge and their class of the highest probability, respectively. For comparison, manually labeled results are also demonstrated at the last row. Generally speaking, the method has good performance with a majority of objects, especially large-scale ones, such as buildings and trees. Compared with the results of contour-based segmentation, over-segmentation in Fig. 24 is reduced greatly, and the problem of under-segmentation (as Failure B in Fig. 17) is solved in this result (highlighted by A in Fig. 24). However, classification errors still exist, where B indicates two cars that are misclassified as buildings, C is trees that are also misclassified as buildings, and D is clumps of bushes, however misclassified as cars. Classification accuracy will be improved through future work.

Contextual Map of Dynamic Environment

As discussed previously, a 3D coordinate can be retrieved for each range pixel, and a range image can be converted into a 3D representation using laser points. Figure 27 shows a map representation of an environment using the geo-reference 3D laser points of L4 and L5, where colors denote different sensor data. In such a map, laser points represent object geometry but do not suggest any contextual knowledge of the objects. Through segmentation and classification, range segments, i.e., laser point clouds, are extracted that represent the data of the same objects, and they are annotated by certain object classes. After segmentation and classification, the laser points of Fig. 26 are annotated and shown in Fig. 27. Here, the color legend is the same as that of Figs. 24 and 25. The result is compared with a manually labeled one, as shown in Fig. 28, where those laser points (i.e., range pixels) that are difficult in



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 26
A map of the environment represented by unlabeled 3D laser points of L4 and L5 [41]

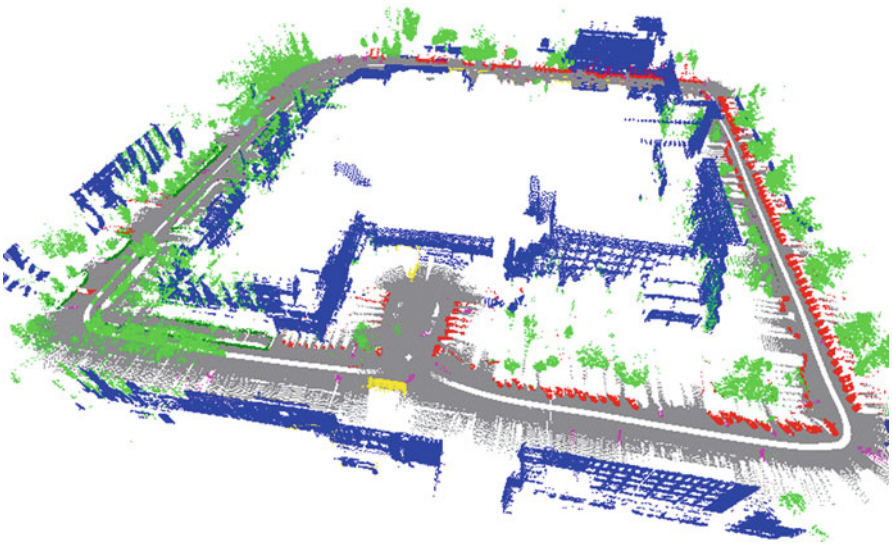


Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 27
A map of the environment represented by annotated 3D laser points of L4 and L5 through simultaneous segmentation and classification [41]

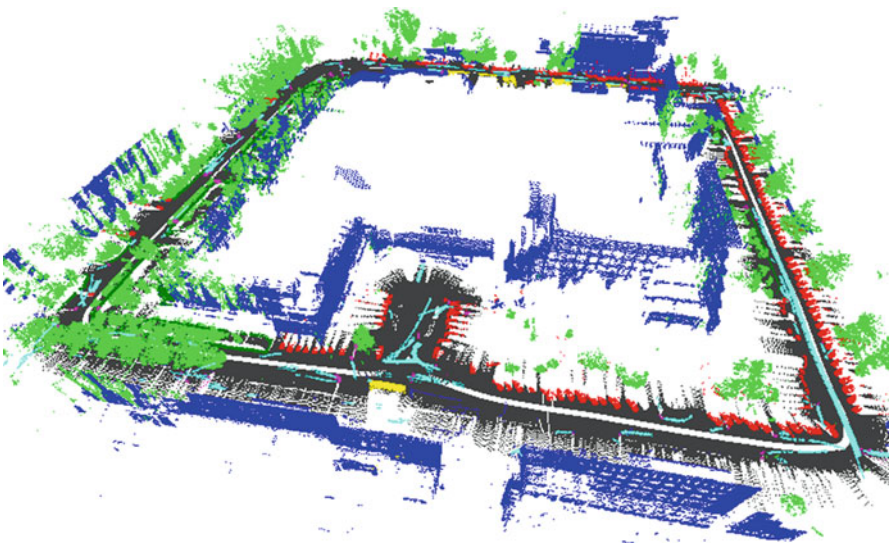
manual discrimination (as the black points in Figs. 24 and 25) are not shown.

In geo-referencing laser measurements of Figs. 26 and 28, the vehicle position of Fig. 5 is used, which is an output of module 2 –SLAMMODT by combinatory exploration of GPS/IMU inputs and a horizontal laser scanner L1. Except for vehicle position, a set of motion trajectories is also obtained as the output of module 2,

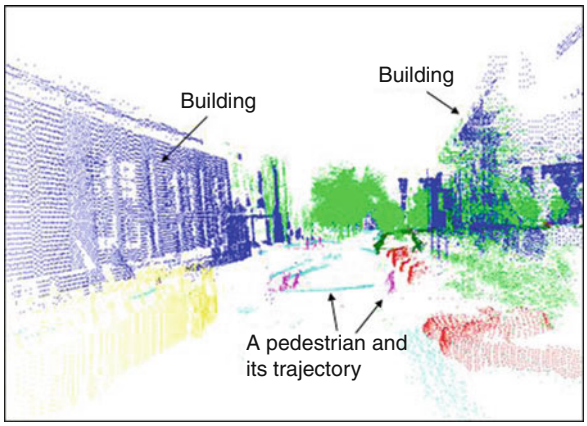
which represents the motion of moving objects at a horizontal level that are captured by L1. With the result of Figs. 27 and 28, a 3D map representation of the potentially dynamic objects, such as cars and pedestrians, as well as those of static ones in the environment are generated. The motion trajectories from module 2 are the only direct measurements of the system to the dynamic entities at the environment, however, they are restricted within a horizontal plane. By fusing both of the data as shown in Fig. 29, where the motion trajectories are overlapped onto the 3D map as shown in light blue points, those true dynamic entities at the moment can be inferred. For example, there are almost no light blue points associated with the parked cars in



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 28
A map of the environment represented by annotated 3D laser points of L4 and L5 through manual labeling [41]



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 29
A map of the environment with the representation of both static and dynamic entities



Dynamic Environment Sensing Using an Intelligent Vehicle. Figure 30
Rendered from a viewpoint on street

Fig. 29, which reflect that the potentially moving objects are static at the moment. A map is rendered in Fig. 30 from a street viewpoint, where a pedestrian as well as its motion is highlighted. The 3D shape of the pedestrian is vividly represented in the 3D map. A long, light blue tail can be associated with the pedestrian, which reflects its walking path. In addition, there is a building entrance nearby where the pedestrian is heading, suggesting a possible reason for such a motion.

Future Directions

In order to generate an omni-directional range sensing of a dynamic urban scene using an intelligent vehicle, a system and algorithmic developments for the fundamental issues of multi-laser sensor system calibration, and scene understanding in contextual map generation are presented, as well as experimental examinations. Although there are still open problems left in each of the key issues, this research intends to demonstrate that a geometric and contextual representation of static objects such as buildings, trees, and roads, as well as the motion at the moment of dynamic entities such as people, bicycles, and cars can be obtained using a vehicle robot system.

It is not only that such a representation is helpful in a robot's online perception and decision, but it is also useful in a broad range of potential applications. A city has not only static objects, such as buildings, trees, and roads, but also moving objects, such as people and cars. A study to analyze the interactions between static and dynamic objects is only available when a sensing technique is able to generate a copy of their simultaneous states. Knowledge of the statistical properties of the relationship between static and dynamic objects is important for an advanced online perception, such as surveillance, or offline systems, such as behavior analysis, architectural design, security planning, and so on. This research focused on a sensing system using an intelligent vehicle, where a 3D map representation of both the static and dynamic entities as well as motions in a dynamic environment can be generated. However, how to best make use of such a sensing technology for potential applications needs further research. The data sets that are generated in this research are in [78].

Bibliography

1. International Society of Photogrammetry and Remote Sensing. http://www.isprs.org/
2. Collins R, Hanson A, Riseman E (1994) Site model acquisition under the UMass RADIUS project. In: Proceedings of arpa image understanding workshop, Monterey, CA, pp 351–358
3. Gruen A (1998) TOBAGO – a semi-automated approach for the generation of 3-D building models. ISPRS J Photogramm Remote Sens 53(2):108–118
4. Forstner W (1999) 3D-city models: automatic and semiautomatic acquisition methods. In: Proceedings photogrammetric week, University of Stuttgart, Institute for photogrammetry, pp 291–303
5. Shiode N (2001) 3D urban model: recent developments in the digital modeling of urban environments in three-dimensions. GeoJournal 52(3):263–269
6. Ellum C, El-Sheimy N (2000) The development of a backpack mobile mapping system. Int Arch Photogramm Remote Sensing XXXII(B2):184–191, Amsterdam
7. He G, Orvets G (2000) Capturing road network data using mobile mapping technology. Int Arch Photogramm Remote Sensing XXXIII(B2):272–277, Amsterdam
8. Silva JFC, Camargo PO, Oliveira RA (2000) A street map built by a mobile mapping system. Int Arch Photogramm Remote Sensing XXXIII(B2):510–517, Amsterdam
9. El-Sheimy N (2000) Mobile multi-sensor systems: the new trend in mapping and GIS applications. IAG J Geodesy, vol. 121, Geodesy beyond 2000: the challenges of the first decade. Springer, Berlin/Heidelberg, pp 319–324
10. Li R (1997) Mobile mapping: an emerging technology for spatial data acquisition. Photogramm Eng Remote Sens 63(9):1085–1092

11. Früh C, Zakhor A (2004) An automated method for large-scale, ground-based city model acquisition. Int J Comput Vis 60(1):5–24
12. Ikeuchi K, Sakauchi M, Kawasaki H, Sato I (2004) Constructing virtual cities by using panoramic images. Int J Comput Vis 53(3):237–247
13. Zhao H, Shibasaki R (2003) Special issue on computer vision system: reconstructing textured CAD model of urban environment using vehicle-borne laser range scanners and line cameras. Mach Vis Appl 14(1):35–41
14. Google Earth (2004) http://earth.google.com
15. Microsoft Virtual Earth (2006) http://www.microsoft.com/virtualearth
16. Google StreetView (2007) http://maps.google.com/help/maps/streetview
17. StreetMapper (2007) http://www.streetmapper.net
18. City Grid (2006) http://www.cybercity.tv
19. Cyber City (2007) http://www.cybercity.tv
20. Hu J, You S, Neumann U (2003) Approaches to large-scale urban modeling. IEEE Comput Graph Appl 23(6):62–69
21. Thrun S (2002) Robotic mapping: a survey. CMU-CS-02-11
22. DARPA (2004) DARPA grand challenge rulebook. http://www.darpa.mil/grandchallenge05/Rules_8oct04.pdf
23. DARPA (2006) DARPA grand challenge rulebook. http://www.darpa.mil/grandchallenge/docs/Urban_Challenge_Rules_121106.pdf
24. Journal of Field Robotics: Special issue on the 2007 DARPA urban challenge, Part I 25(8)
25. Journal of Field Robotics: Special issue on the 2007 DARPA urban challenge, Part II 25(9)
26. Nuchter A, Lingemann K, Hertzberg J, Surmann H (2007) 6D SLAM – 3D mapping outdoor environments. J Field Robot 24(8/9):699–722
27. Zhao H, Shibasaki R (2003) A vehicle-borne urban 3D acquisition system using single-row laser range scanners. IEEE Trans SMC Part B: Cybern 33–4:658–666
28. Allen P, Atamos I, Gueorguiev A, Gold E, Blaer P (2001) AVENUE: automated site modeling in urban environments. In: Proceedings of the 3rd international conference on 3D digital imaging and modeling, Quebec City, pp 357–364
29. Georgiev A, Allen PK (2004) Localization methods for a mobile robot in urban environments. IEEE Trans Robot Automat (TRO) 20(5):851–864
30. Hahnel D, Burgard W, Fox D, Thrun S (2003) An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems, Las Vegas, pp 206–211
31. Wang CC (2004) Simultaneous localization, mapping and moving object tracking. PhD dissertation, Carnegie Mellon University, CMU-RI-TR-04-23
32. Vu T, Aycard O, Appenrodt N (2007) Online localization and mapping with moving object tracking in dynamic outdoor environment. In: Proceedings IEEE intelligent vehicle symposium, Istanbul, pp 190–195
33. Weiss T, Schiele B, Dietmayer K (2007) Robust driving path detection in urban and highway scenarios using a laser scanner and online occupancy grids. In: Proceedings of the IEEE intelligent vehicle symposium, Istanbul, pp 184–189
34. Zhao H, Chiba M, Shibasaki R, Shao X, Cui J, Zha H (2008) SLAM in a dynamic large outdoor environment using a laser scanner. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), Pasadena, CA, pp 1455–1462
35. Velodyne HDL-64E (2007) http://www.velodyne.com/lidar/products/overview.aspx
36. SICK LMS2** (2001) http://www.sick.com
37. Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping: Part I. IEEE Rob Autom Mag 13(2):99–108
38. Bailey T, Durrant-Whyte H (2006) Simultaneous localization and mapping: Part II. IEEE Rob Autom Mag 13(3):108–117
39. Streller D, Dietmayer K (2004) Object tracking and classification using a multiple hypothesis approach. In: Proceedings of the IEEE intelligent vehicles symposium, Parma, pp 808–812
40. Fayad F, Cherfaoui V (2007) Tracking objects using a laser scanner in driving situation based on modeling target shape. In: Proceedings of the IEEE intelligent vehicle symposium, Istanbul, pp 44–49
41. Zhao H, Liu Y, Zhu X, Zhao Y, Zha H (2010) Scene understanding in a large dynamic environment through a laser-based sensing.In: IEEE international conference on robotics and automation, Anchorage, pp 127–133
42. Zhao H, Xiong L, Jiao Z, Cui J, Zha H (2009) Sensor alignment towards an omni-directional measurement using an intelligent vehicle. In: Proceedings of the IEEE intelligent vehicle symposium, Kobe, pp 292–298
43. Zhao H, Shibasaki R (2001) High accurate positioning and mapping in urban area using laser range scanner. In: Proceedings of IEEE intelligent vehicles symposium, Tokyo, pp 125–132
44. Zhao H, Chiba M, Shibasaki R, Shao X, Cui J, Zha H (2009) A laser scanner based approach towards driving safety and traffic data collection. IEEE Trans Intell Transport Syst 10(3):534–546
45. Cheok GS, Leigh S, Rukhin A (2002) Calibration experiments of a laser scanners. NISTIR 6922:121
46. Santala J, Joala V (2003) On the calibration of a ground-based laser scanner. TS12.4, FIG working week
47. Mahlisch M, Hering R, Ritter W, Dietmayer K (2006) Heterogeneous fusion of video, Lidar, ESP data for automotive ACC vehicle tracking. In: Proceedings of IEEE international conference on multisensor fusion and integration for intelligent systems, Heidelberg, pp 139–144
48. Rodriguez F, Sergio A, Fremont V, Bonnifait P (2008) Extrinsic calibration between a multi-layer LIDAR and a camera. In: Proceedings of IEEE international conference on multisensor fusion and integration for intelligent systems, Seoul, pp 214–219
49. Zhao H, Chen Y, Shibasaki R (2007) An efficient extrinsic calibration of a multiple laser scanners and cameras' sensor system on a mobile platform. In: IEEE intelligent vehicles symposium, Istanbul, pp 422–427

50. Gao C, Spletzer JR (2010) On-line calibration of multiple LIDARs on a mobile vehicle platform. In: Proceedings of IEEE international conference on robotics and automation, Anchorage, pp 279–284
51. Besl PJ, McKay ND (1992) A method for registration of 3-D shape. IEEE Trans Pattern Anal Mach Intell 14:239–256
52. Chen Y, Medion G (1992) Object modeling by registration of multiple range images. Image Vis Comput 10(3):145–155
53. Zhang Z (1994) Iterative point matching for registration of a free-from curves and surfaces. Int J Comput Vis 13:119–152
54. Hahnel D, Burgard W, Thrun S (2003) Learning compact 3d models of indoor and outdoor environments with a mobile robot. Rob Autom Syst 44:15–27
55. Althaus P, Christensen H (2003) Behavior coordination in structured environments. Adv Robot 17(7):657–674
56. Mendes A, Nunes U (2004) Situation-based multi-target detection and tracking with laserscanner in outdoor semi-structured environment. In: Proceedings IEEE/RSJ international conference on intelligent robots and systems, Sendai
57. Posner I, Cummins M, Newman P (2008) Online generation of scene descriptions in urban environments. Rob Autom Syst 56(11):901–914
58. Douillard B (2009) Vision and laser based classification in urban environments. PhD thesis, University of Sydney
59. Nuchter A, Lingemann K, Hertzberg J, Surmann H (2009) 6D SLAM – 3D mapping outdoor environments. J Field Robot 24(8–9):699–722
60. Martinez-Mozos O, Stachniss C, Burgard W (2005) Supervised learning of places from range data using AdaBoost. In: IEEE international conference on robotics and automation (ICRA), Barcelona, pp 1742–1747
61. Triebel R, Kersting K, Burgard W (2006) Robust 3D scan point classification using associative Markov networks. In: IEEE international conference on robotics and automation (ICRA), Orlando, FL
62. Douillard B, Fox D, Ramos FT (2008) Laser and vision based outdoor object mapping. In: Proceedings of robotics: science and systems, Zurich
63. Posner I, Cummins M, Newman P (2009) A generative framework for fast urban labeling using spatial and temporal context. Auton Robot 26(2–3):153–170
64. Hoover A, Jean-Baptiste G, Jiang X, Flynn PJ, Bunke H, Goldgof D, Bowyer K (1996) A comparison of range image segmentation algorithms. IEEE Trans Pattern Anal Mach Intell 18(7):673–689
65. Katsoulas D, Bastidas CC, Kosmopoulos D (2008) Superquadic segmentation in range image via fusion of region and boundary information. IEEE Trans Pattern Anal Mach Intell 20(5):781–795, 2008
66. Weingarten J, Siegwart R (2005) EKF-based 3D SLAM for structured environment. In: IEEE/RSJ international conference on intelligent robots and systems, Edmonton, Alberta, pp 2089–2094
67. Han F, Tu Z, Zhu SC (2004) Range image segmentation by an effective jump-diffusion method. IEEE Trans Pattern Anal Mach Intell 26(9):1138–1153
68. Borenstein E, Ullman S (2008) Combined top-down/bottom-up segmentation. IEEE Trans Pattern Anal Mach Intell 30(12):2109–2125
69. Malisiewicz T, Efros AA (2008) Recognition by association via learning per-exemplar distances. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Anchorage
70. Porway J, Wange K, Zhu SC (2008) A hierarchical and contextual model for aerial image understanding. In: Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR), Anchorage, pp 1–8
71. Tu Z, Chen X, Yuille A, Zhu SC (2005) Image parsing: unifying segmentation, detection and recognition. Int J Comput Vis 63(2):113–140
72. Golovinskiy A, Kim VG, Funkhouser T (2009) Shape-based recognition of 3D point clouds in urban environments. In: IEEE international conference on computer vision, Kyoto, pp 2154–2161
73. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22:888–905
74. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. Int J Comput Vis 59:167–181
75. Jiang X, Bunke H (1994) Fast segmentation of range images into planar regions by scan line grouping. Mach Vis Appl 7:115–122
76. Rosin PL, West GAW (1995) Nonparametric segmentation of curves into various representations. IEEE Trans Pattern Anal Mach Intell 17:1140–1153. http://users.cs.cf.ac.uk/Paul.Rosin
77. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/∼cjlin/libsvm
78. PKU Omni Smart Sensing – POSS (2010) http://www.poss.pku.edu.cn